NEW PHARMACOLOGICAL TARGETS FOR CYSTIC FIBROSIS TREATMENT FROM OMICS PROFILING OF F508del-CFTR EXPRESSING CELLS.

Ph.D. student: Clarissa Braccia

Dottorato in Scienze e Tecnologie della Chimica e dei Materiali

Curriculum: Drug Discovery and Nanobiotechnologies

Università degli studi di Genova - Istituto Italiano di Tecnologia (D3-PharmaChemistry)

Advisors:

Armirotti Andrea (IIT)

Bandiera Tiziano (IIT)

Caviglioli Gabriele (DIFAR, Unige)

Contents

1 Introduction

1.1 Cystic Fibrosis

1.1.1 Pathophysiology of the disease

Cystic Fibrosis (CF) is a lethal, inherited and autosomal recessive disorder that affects approximately 1 every 3000 births [1].

CF is caused by mutations in the gene encoding for the cystic fibrosis transmembrane conductance regulator (CFTR) protein, a transmembrane anion channel situated at the apical membrane of epithelial cells of the airway, the gastrointestinal tract, the pancreas, and the biliary and sweat ducts. CFTR regulates the transport of chloride and bicarbonate ions in and out of the cells. This protein has an important role in the regulation of epithelial ion and water transport and fluid homeostasis.

The first comprehensive description of CF was given in 1938 by Dr. Dorothy H. Andersen [2], a pathologist at the New York Babies Hospital, who introduced the term "Cystic Fibrosis of the pancreas" to describe the destruction of the pancreatic exocrine function as a result of the disease. Another major advance in the history of CF research was achieved by Dr. Paul Di Sant'Agnese in 1953, who demonstrated that the sweat of CF children had an excessive salt content [3]. His findings led to the first method for CF diagnosis, the sweat test, which measures the concentration of sodium and chloride in the sweat. In the 1980s, Dr. Paul M. Quinton defined the salt transport abnormality and demonstrated the attenuated chloride transport in sweat ducts [4], [5]. A similar abnormality in the respiratory epithelium was also described by Dr. Michael Knowles [6] in 1983.

The CFTR gene, located on the long arm of chromosome 7 at position q31.2 [7], was identified and cloned in 1989 [8]. Since its discovery, both the CFTR structure and function have been deeply investigated. These efforts improved our understanding of CF pathophysiology but also outlined the complexity of the disease.

CFTR is 1480-aminoacids glycoprotein belonging to the adenosine triphosphate (ATP)-binding cassette (ABC) transporter superfamily, and it is composed of five domains [9]:

- two transmembrane domains forming the channel pore (TMD1 and TMD2), each with six hydrophobic α-helix transmembrane segments (TMs);
- two cytosolic nucleotide-binding domain (NBD1 and NBD2) involved in the ATP binding;
- a unique regulatory domain (RD) with multiple protein kinase A (PKA)-dependent phosphorylation sites, that control CFTR activity as an ATP-gated ion channel.

Moreover, CFTR presents two extracellular asparagine (Asn)-linked N-glycosylation sites between the transmembrane segments TM7 and TM8 [10]. The schematic structure of CFTR is shown in Figure 1.

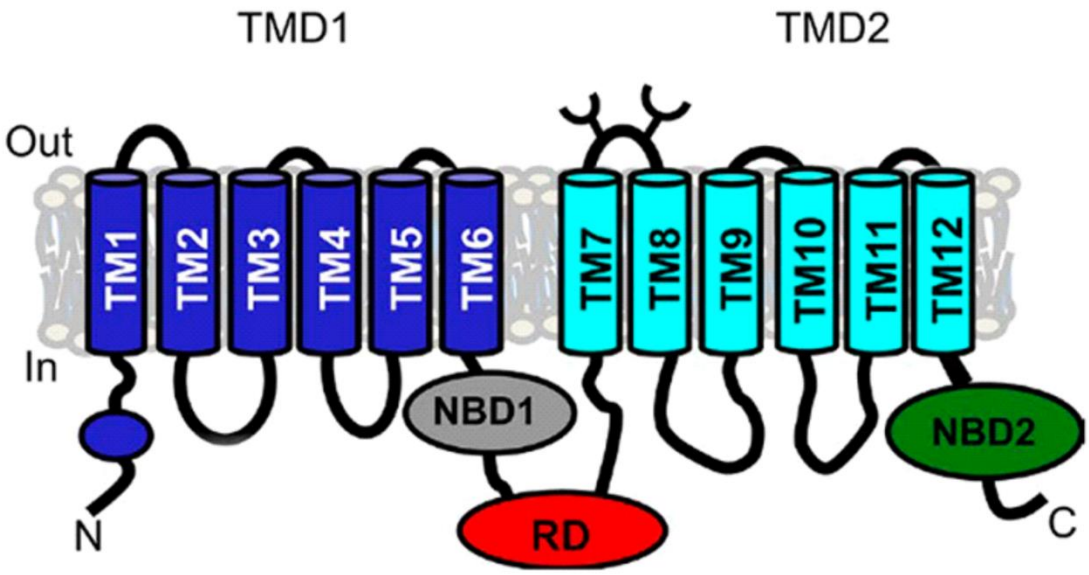


Figure 1: Schematic structure of CFTR channel. Figure adapted from Hwang TC: Structural mechanisms of CFTR function and dysfunction [9].

The "gating" of the channel is a finely regulated process: the activation starts with the phosphorylation of R domain through the protein kinase A (PKA) and the recruitment of ATP to the NBD1 and NBD2, which dimerize to open the channel. The subsequent ATP hydrolysis causes the NBDs dissociation to close the channel [11].

CFTR is a cyclic adenosine monophosphate (cAMP)-regulated anion channel: in cells, high levels of cAMP activate the PKA that phosphorylates the R domain of CFTR to regulate the channel activity [12]. CFTR transport Cl^- and HCO_3^- across the epithelium in the respiratory tract, pancreas, gastrointestinal tract, the biliary and sweat ducts and part of the reproductive organs [9].

Mutations in CFTR protein result in dysfunctional ion transport across the apical membrane at the surface of several epithelia, generating thickened and dehydrated secretions. In the lung, this leads to a decrease in the mucociliary clearance, favoring bacterial colonization and progressive obstruction of the ducts, as shown in Figure 2 [13].

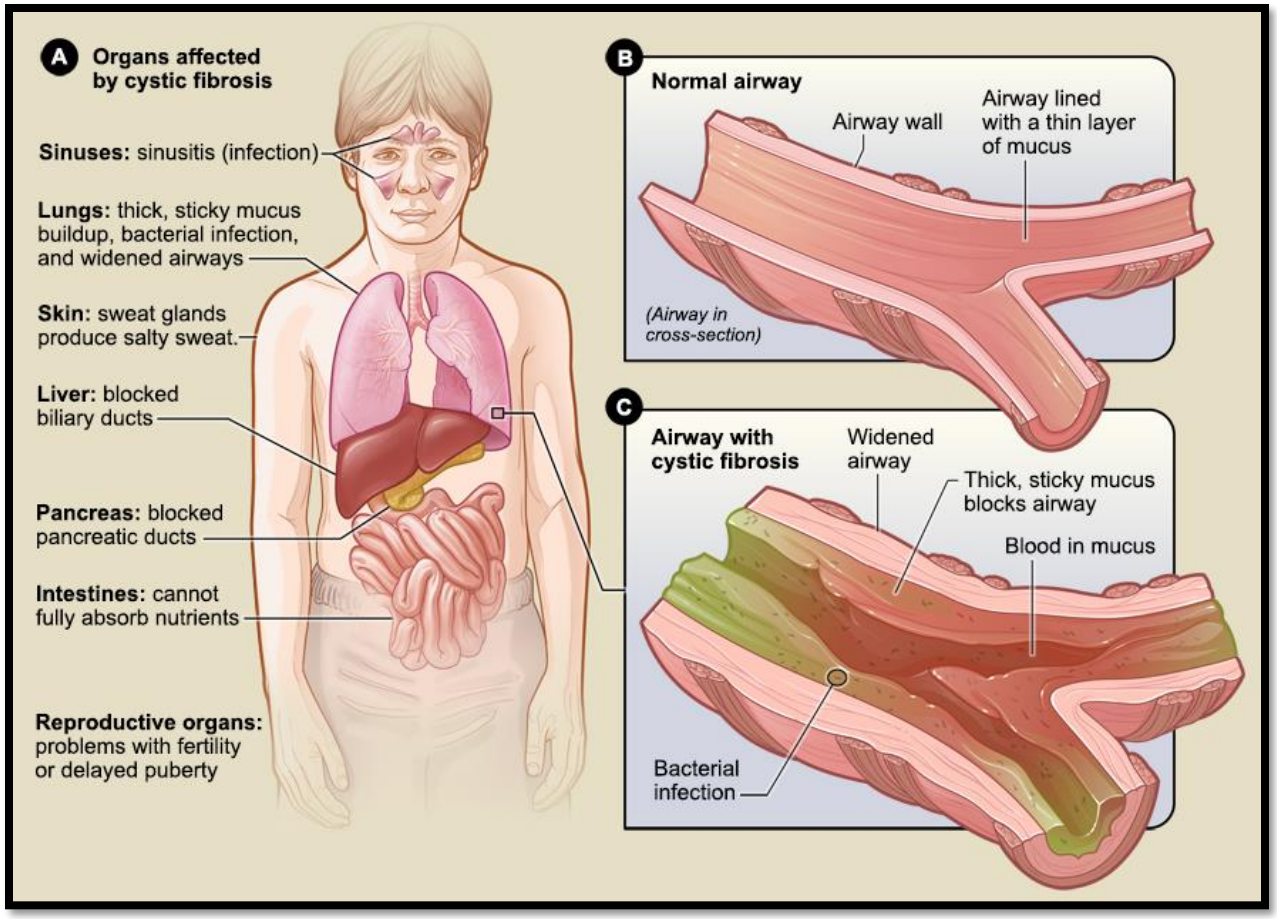


Figure 2: Panel A shows the organs that cystic fibrosis can affect. Panel B shows a cross-section of a normal airway. Panel C shows a cross-section of the airway of a CF patient. Figure adapted from National Heart Lung and Blood Institute (NIH) website.

The dysfunctional ion transport across the apical membrane severely impacts the organs where CFTR is expressed with the respiratory system being the worst affected. The impaired activity of CFTR in the airway epithelium causes dysregulation of the volume and the composition of the airway surface liquid (ASL). Indeed, the airway epithelial ion transport is essential to generate osmotic gradients that provide the driving force for transepithelial fluid movement [14]. The ASL volume is finely regulated by the airway epithelium and its impaired work can lead to ASL depletion, promoting mucus adhesion and chronic bacterial colonization in the lungs, the major cause of morbidity and mortality [15].

1.1.2 Classes of CFTR mutations

Although over 2000 CFTR variants have been identified so far, pathogenicity has been demonstrated only for approximately 350 variants [16]. Only a relatively limited number of all the CFTR mutations are known to be related to CFTR channel dysfunction, whereas the functional consequences of some rare sequence mutations remain unknown. The CF-causing mutations have been traditionally classified in six classes [17], according to their final effect on CFTR (a summary of the classification is shown in Figure 3):

- ***Class I***: failure in protein synthesis due to frameshift or nonsense mutations; for example, G542X is a nonsense or stop mutation, where the introduction of a premature termination codon (or stop codon) results in premature cessation of translation and production of truncated CFTR protein.
- ***Class II***: unstable conformation and early degradation of the protein; this class includes F508del mutation and the resulting protein shows folding or maturation defects.
- ***Class III***: low transport of chloride through the channel due to limited channel gating caused by ineffectual binding of nucleotide; an example is G551D, which accounts for 2–3% of CFTR mutations worldwide.
- ***Class IV***: low ions conductance through the channel due to conductance defects, even though the channel is able to open and close.
- ***Class V***: small amount of functional and normal protein because of limited transcriptional regulation.
- ***Class VI***: destabilization at the apical membrane, leading to high turnover of CFTR at the cell surface.

In 2016, De Boeck and Amaral divided the first class mutations into Class I (carrying stop-codons) and Class VII (which lacks mRNA transcription) [18]. Class VII mutations are considered among the most severe CFTR mutations: even though Class VII has the same outcome as Class I, that is the absence of CFTR protein, it cannot be rescued by corrective therapy.

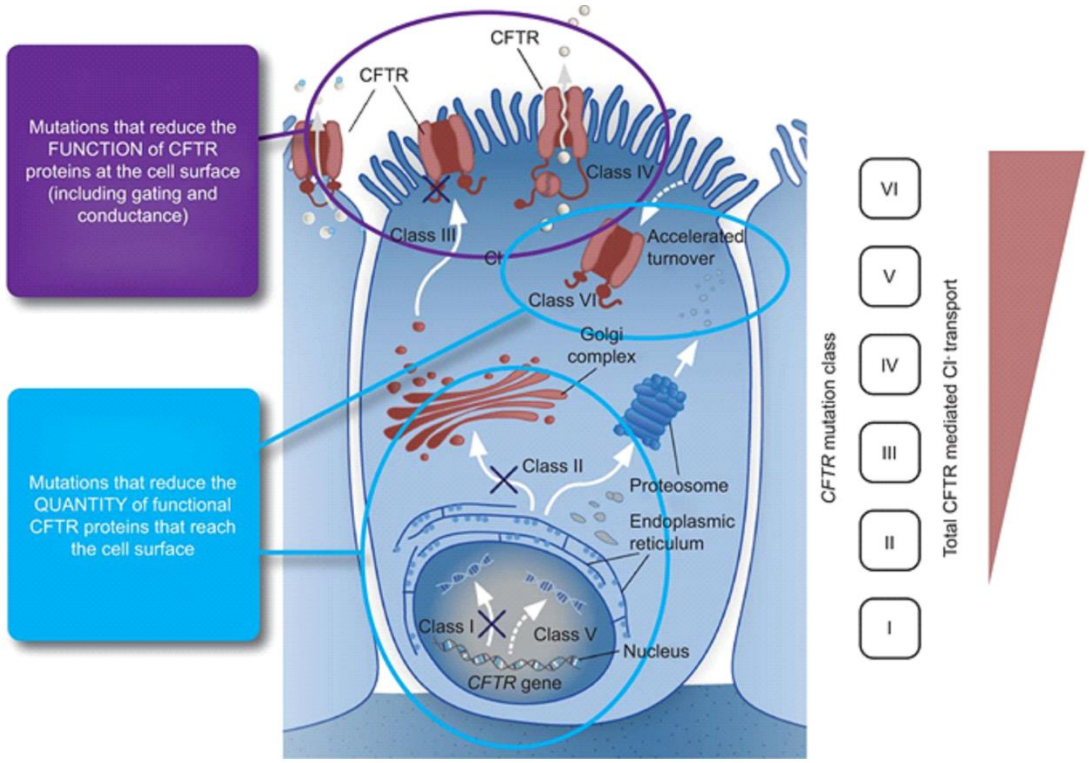


Figure 3: Categorization of CFTR mutations in six classes. Class I, II, V and VI are responsible for a reduction of functional CFTR that reaches the membrane; class III and IV show a reduction of CFTR function. Picture taken from Derichs N., Targeting a genetic defect: cystic fibrosis transmembrane conductance regulator modulators in cystic fibrosis [17].

Only four CFTR mutations show a frequency of 1-3% among CF patients: G551D, W1282X, G542X and N1303K. The most common mutation is a deletion of the phenylalanine in position 508 (F508del), which shows at least one allelic frequency of around 90% among CF patients [19]. The F508del-CFTR is incorrectly folded, causing retention at the endoplasmatic reticulum (ER) and subsequent proteasomal degradation. Moreover, the small amount of mutant CFTR that reaches the membrane is unstable and shows gating defects [20].

1.1.3 Synthesis and Trafficking of CFTR

CFTR biogenesis, as an integral transmembrane glycoprotein, begins in the endoplasmic reticulum (ER). The newly synthesized CFTR polypeptide chain is co-translated into the ER membrane and when it emerges into the ER lumen, CFTR is N-glycosylated at the Asn894 and Asn900 residues, which are located in the fourth extracellular loop [21].

The N-glycosylation plays an essential role in glycoproteins intracellular processes, such as folding and trafficking [22], since the sugars moieties act as stabilizers and recognition targets. For CFTR, the glycosylation is critical for the quality control processes in the ER that checks the correct folding status of those proteins that are sent to the plasma membrane (PM) through the Golgi complex. The glycosylated immature form of wild type (wt)-CFTR can traffic to the Golgi complex, where it is processed by several Golgi glycosyltransferases to create the fully mature form of wt-CFTR [23].

Cells have different quality control mechanisms that recognize the misfolded proteins and then target them for degradation. For CFTR, cells have several levels at which this happens [24].

In order to exit from ER and reach the Golgi, CFTR must undergo four ER quality control (ERQC) checkpoints [25], summarized in Figure 4:

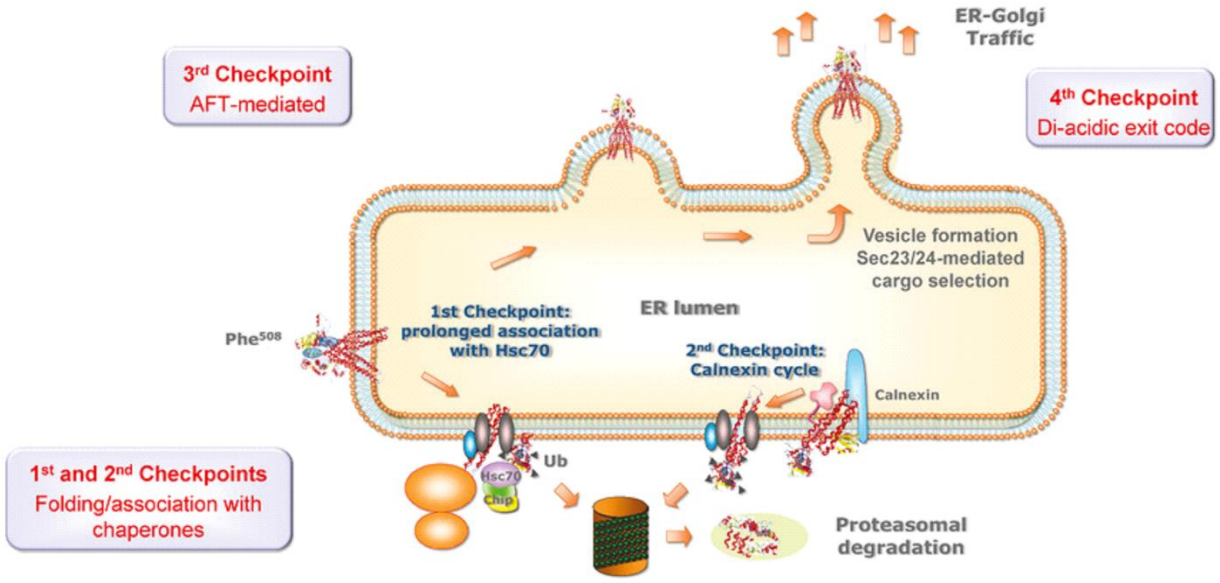


Figure 4: Four checkpoints for the ER quality control of CFTR. Figure taken from Farinha C. M., Control of cystic fibrosis transmembrane conductance regulator membrane trafficking: not just from the endoplasmic reticulum to the Golgi [23].

1) The first checkpoint is represented by the ***chaperone complexes containing Hsc70 (Heat-shock cognate)/Hsp70 and Hsp90 (Heat-shock protein 90)*** [26], able to bind the nascent CFTR in order to assess its conformation. This is the major mechanism for the trapping of F508del-CFTR; in fact, Hsp/Hsc70 is able to recognize (and strongly bind to) the exposed hydrophobic residues of the mutant protein, given its failure to acquire the proper conformation.
2) The second ERQC checkpoint involves the ***glycosylation of CFTR and the calnexin cycle*** [27]. Calnexin, able to bind N-glycosyl residues, is a key regulator of the ER quality control system and its prolonged association with F508del-CFTR promotes the degradation of the mutant protein.

3) The third checkpoint occurs at the ***ER exit sites***, where misfolded CFTR is retained because it exposes the ER retention motifs – four arginine-framed tripeptides (AFTs) [28].
4) The fourth checkpoint is a ***positive export signal*** (the di-acidic exit code DAD, located in the NBD1) that mutant CFTR fails to expose to exit ER [29].

Proteins which are recognized as unfolded or misfolded by quality control mechanisms of the cell (F508del-CFTR *e.g.*) are almost completely retained in the ER and then targeted for premature degradation through the ubiquitin-proteasome pathway (UPP) [30].

Besides this conventional trafficking pathway, a "nonconventional" trafficking route from ER to the Golgi apparatus was described for CFTR [31], even though it has limited therapeutic significance. This route is insensitive to the checkpoints of the conventional route and exploits tubular structures migrating peripherally to the central Golgi [32]. With this route, the COPII (coated proteins type II) machinery takes both the folded and misfolded CFTR but only the properly folded protein will reach the Golgi. Through the COPI (coated protein I) vesicles, the misfolded protein is recycled in the ER, where the protein can repeat the folding process and then can be sent either to the Golgi or to the proteasome for degradation.

The correct folding of the protein is essential not only for the ERQC machinery, but also for the CFTR retention at the plasma membrane.

The levels of CFTR at the PM are regulated by three processes: anterograde trafficking (from the trans-Golgi network to the PM), endocytosis in clathrin-coated vesicles, and recycling of the endocytosed CFTR [33]. Of these three processes, endocytosis and recycling are extensively regulated [34]: indeed, several studies show that the internalized CFTR is stored in early endosomes, which are used for the continuous recycling of CFTR [35], [36] from endosomes to the PM.

Since the endocytosis of CFTR is a rapid process, if compared to its biosynthesis and maturation [37], the recycling of internalized CFTR is crucial to maintain a functional pool of anion channel at the membrane surface [38].

The CFTR trafficking, as well as its turnover and retention at the surface of epithelial cells, is a complex process in which the folding status of the protein is assessed [39].

Rab GTPaes, one of the largest branches of the Ras superfamily [40], are key regulators of the intracellular protein transport [41] and they are also involved in the delivery of CFTR to the plasma membrane, as shown in Figure 5.

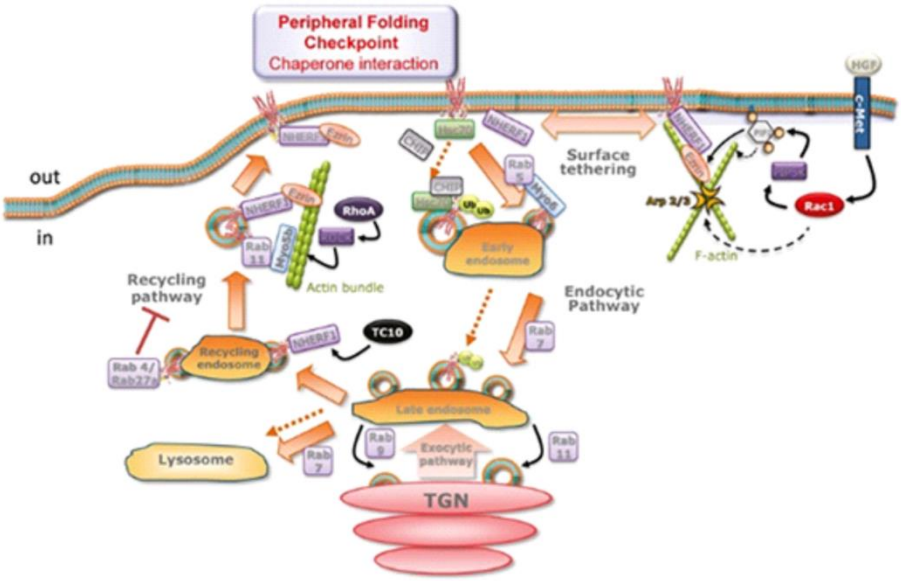


Figure 5: Regulation of CFTR trafficking and retention at the cell surface by Ras GTPases. Picture taken from Farinha C. M., Control of cystic fibrosis transmembrane conductance regulator membrane trafficking: not just from the endoplasmic reticulum to the Golgi [23].

Mature CFTR exits the trans-Golgi network (TGN) in vesicles belonging to the intracellular protein transport to reach the plasma membrane [42]. CFTR insertion into the plasma membrane is driven by the myosin myo5b, Rab 11 and the PDZ adaptor NHERF1 [43].

At the PM, CFTR must undergo the final checkpoint: if CFTR passes this last checkpoint, it can be tethered to the plasma membrane through the actin cytoskeleton. Otherwise, the misfolded CFTR is rapidly ubiquitinated and caught by the endocytic pathway [44]. At this point CFTR can be either deubiquitinated, refolded and recycled to the PM or sent to the lysosome via Rab7 for its degradation [44]. Interestingly, Rab9 and Rab11 can lead CFTR back to the TGN in the sorting endosome. On the contrary, if CFTR binds to Rab4 and Rab27, it is retained at the intracellular endosomal compartment, preventing its trafficking to the PM.

1.1.4 Drugs for CF

In the 1980s, children with cystic fibrosis rarely lived beyond adolescence. Over the last decade, life expectancy increased, thanks to the use of antibiotics, anti-inflammatory and mucolytic drugs as symptomatic therapies. Nevertheless, these therapies only treat the manifestations of the disease, without correcting the channel defect [45].

At present, the disease-specific therapeutic options for CF patients depend on the type of mutation in the CFTR gene. Two classes of drugs are available [46] exploiting different mechanisms of action, summarized in Figure 6:

- ***Potentiators:*** molecules that activate the mutant CFTR channel present at the plasma membrane. The only approved potentiator, the molecule VX-770 (trade name: Ivacaftor), can be used to treat patients having the G551D and other mutations causing gating defect.
- ***Correctors:*** molecules that increase the trafficking of the mutant CFTR channel to the plasma membrane. These classes of drugs are used for patients expressing the F508del mutation (around 80% of CF patients). Patients homozygous for the F508del mutation are treated with Orkambi, a combination of the corrector Lumacaftor (trade name for the molecule VX-809) and the potentiator Ivacaftor.

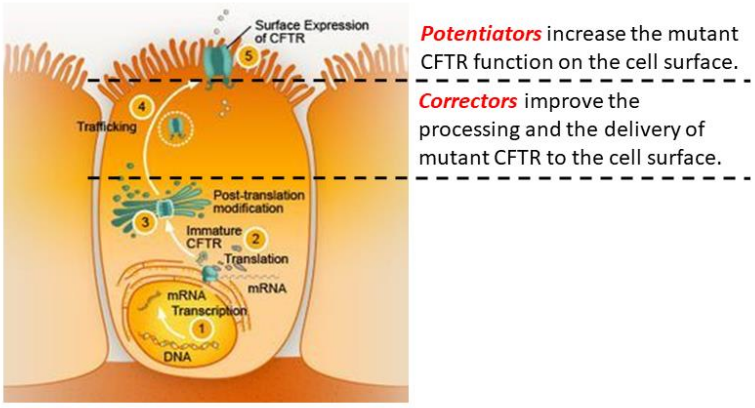


Figure 6: Site and mechanisms of action of the two classes of CFTR modulator drugs. Figure adapted from De Boeck K. Cystic fibrosis in the year 2020: A disease with a new face [47].

It has been demonstrated that a combination therapy with molecules exploiting different mechanisms of action is required to achieve a therapeutically relevant functional rescue of mutant CFTR [17]. For this reason, other combination drugs were developed, such as the drug Symdeko, the combination of the corrector Tezacaftor (also known as VX-661) with Ivacaftor for the treatment of patients having the F508del mutation on one allele and a so-called “residual function” mutation on the second allele.

Ivacaftor produces a substantial improvement in the quality of life of CF patients. To the contrary, patients receiving Orkambi or Symdeko alone have a modest improvement in their conditions [48]. The new Kaftrio [49], a combination of three drugs (the two correctors Elexacaftor and Tezacaftor together with the potentiator Ivacaftor), aims at radically improving the quality of life of CF patients.

1.1.5 Models to study CF

To deeply understand the molecular biology of diseases, cellular and animal models that represent the disease processes are needed. In recent years, advances in genetic engineering technologies helped the development of sophisticated models to investigate human diseases such as neurodegenerative conditions, cancers, and other genetic disorders. To choose the correct model, researchers need to know if the model system reflects the human response. Indeed, researchers must work with models that most closely resemble the human system. For CF, finding the correct model is complicated due to the multi-organ nature of the disease [50].

Modern genome editing technologies allow the researchers to build a reliable model, inserting a human gene into an animal species relevant to study new therapies for human disease. These innovations enable scientists to reveal more about the pathogenesis of a certain disease. Nevertheless, the genetic modification can have undesirable effects causing, for example, the death of the animal. For this reason, it is extremely important to validate the genetics and the phenotype when considering a new model to study a biological function or disease.

Today scientists know that, even if a perfect *in-vivo* model for CF is not available yet, they can compare studies on animals of different species [51] to understand this complex disease. Because of the difficulties in having a well-established *in-vivo* model, many studies are conducted with immortalized cell line models of CF and the validation is then performed in animal models or primary CF patient tissues.

1.1.5.1 In-vitro models

The *in-vitro* models can be divided into two classes: the primary cells and the immortalized cells.

The ***primary cells*** directly derive from excised organs. To study CF, primary human airway epithelial are generally used. The advantage of their use is that these type of cells maintain the endogenous expression of CFTR, enabling the researchers to study CFTR mutation in a native environment. However, they are quite difficult to establish and maintain and can be used only for a limited period [52].

On the other hand, the ***immortalized cells*** guarantee higher experimental consistency and can be grown for prolonged periods, obtaining a higher number of cells for further analysis. On the other hand, these cells could be quite different from the *in-vivo* tissue from which they derive.

To study CF, several immortalized cells are commercially available, the most popular being the cystic fibrosis bronchial epithelium 41o- cells (CFBE41o-, shown in Figure 7). The CFBE41o- cells are a CF human bronchial cell line, derived from a CF patient homozygous for F508del-CFTR mutation and immortalized with the origin-of-replication defective SV40 plasmid (psVori-) [53]. These cells show all the ions transport defects caused by transfected mutant CFTR and, when cultured properly, they also form tight junctions to give a polarized epithelium [54].

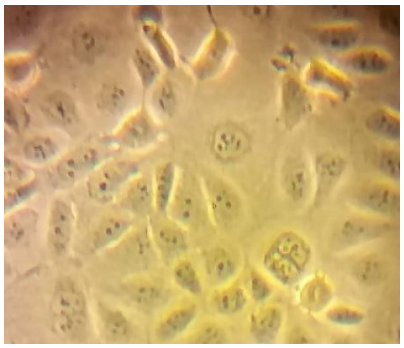


Figure 7: Picture of CFBE41o- in culture.

1.1.5.2 In-vivo models

In-vitro models described above show many advantages since the cells are of human origin and can be easily manipulated. On the contrary, animal models have their own cellular and organ physiology, which can differ from humans.

Nevertheless, CF animal models are essential to understand the pathogenesis and to develop new drugs. The researchers in the CF field have access to multiple species that offer unique advantages, summarized in Figure 8.

Disease	Human	Pig	Ferret	Rabbit	Rat	Mouse
Spontaneous Lung Infections	✔	✔	✔	?	X	X
Impaired Growth	✔	✔	✔	✔	✔	✔
Intestinal	✔	✔	✔	✔	✔	✔
Exocrine Pancreatic	✔	✔	✔	?	X	X
Endocrine Pancreas	✔	✔	✔	?	?	X
Hepatic	✔	✔	✔	?	?	X
Gallbladder	✔	✔	✔	?	NA	X

Figure 8: Organ disease phenotype in the CFTR-knockout animal models. The organs affected in CF patients are listed on the left; for each animal model check-marks indicate the presence of the organ-specific phenotype; question marks are symptoms yet to be demonstrated; X indicates the lack of organ-specific symptoms and NA means not applicable since rats don't have the gallbladder. Figure adapted from Rosen, B. H., Animal and model systems for studying cystic fibrosis [50].

Mouse:

Three years after the discovery of CFTR gene, the first generation of CFTR-knockout mice was engineered [55] and several other models followed soon after. To date, several CF models were described in literature [56] to study the two most common mutations in human, the F508del and G551D mutations. Even though CFTR-knockout mice show an intestinal phenotype, it was clear from the first model developed that they do not show spontaneous lung infections found in CF patients [57]. The hypothesis for the lack of spontaneous lung infection is that the mixed genetic background of the original strain influences the development of lung pathology: for example, unlike humans, mice don't have the submucosal glands throughout the airways, but only in the trachea.

Pig and ferret:

Pigs and ferrets are the only CF *in-vivo* models that show the entire spectrum of CF-patients phenotype. As shown in Figure 8, they not only have inflammatory and infectious lung disease but show also all the CF-related symptoms [58].

However, pigs and ferrets are quite challenging species (they have prolonged gestational period and a long time for sexual maturation) and their maintenance requires considerable resources that significantly limit their widespread use in CF research.

Rat:

Though CFTR-knockout rats are easy to manage as compared to pigs or ferrets, it is still not clear whether they develop spontaneous lung infections or not. In fact, their tracheal surface shows a reduced airway surface liquid layer but their mucociliary clearance is efficient [59].

Rats could represent a valuable resource to study other manifestations of the disease, such as the events that lead to the abnormalities of the male reproductive tract [60] or the dental complications [61], not developed by CF mice models.

Rabbit:

Since the rabbit is the newest model for CF, the phenotype is still poorly characterized. Thus far, rabbits showed only gut obstruction at weaning, making the rabbit suitable to study CF intestinal phenotype.

To introduce the F508del mutation in the CFTR gene in the rabbits, the CRISPR/Cas9-targeting technology was used [62]. This technology has a great potential if applied to the *in-vitro* models of epithelia to study molecular processes impaired by mutant CFTR.

1.2 Omics

Omics are a set of biomolecular disciplines that provide a comprehensive assessment of a class of molecules [63]. Omics studies require a considerable investment of time, skills and money but they also represent the advanced tool to investigate the chemical space composition of a living organism. Omics data, if correctly used, are able to highlight the biological pathways altered by disease. In recent years, the technological advances and development of computational tools made possible the high-throughput analysis of biological molecules. The results of omics studies can be used both for biomarker discovery and for the understanding of biological pathways associated with the disease.

A biological pathway is a sequence of interactions between endogenous compounds in a cell that leads to a certain product. The aim of a biological pathway is to control the activity of the cell and its ability to change its behavior in response to stimuli. The correct regulation of a biological pathway is essential to prevent alterations that can lead to pathological states. For this reason, the study of alterations in biological pathways can be useful for researchers to discover new targets and thus to develop new and more effective drugs for a particular disease.

Genomics was the first omics discipline to be developed. It aims at characterizing the genome, in order to specify the coding and the non-coding sequences, their functions and structures. The project for the sequencing of the human genome, concluded in 2001 [64], led to the identification of more than 20000 protein coding genes.

Through the transcription and the translation processes, the coding genes are translated into proteins. With transcriptomics, it is possible to study and quantify the expression of the transcripts within a cell. The transcriptome of a cell is the complete set of RNA (Ribonucleic acid) species and it comprises both the coding RNA (which will be translated into proteins) and the non-coding RNA (involved in several processes, such as post-transcriptional control) [65].

From a single gene, different protein isoforms can be translated, through splicing processes; moreover, the activity of the proteins can be regulated by post-translational modifications (PTMs), such as phosphorylation or methylation [66]. Given such a relevant added complexity, the exact number of *proteoforms* present in a cell is yet to be discovered [67].

Proteins are the actual effectors of biological processes from the regulation of gene expression to the catalysis of chemical reactions. The central role of proteins in cell biology is highlighted by the fact that diseases often occur when one or more proteins show impaired function and also by the fact that most available drugs target proteins. Other essential molecules present in cells are the metabolites, which are the substrates, intermediates and products of all the biological processes. Metabolites (and their complexity: the metabolome) are the closest molecular representation of the phenotype.

If the genome is static in time, protein and metabolite profiles in a cell continuously change over time, from tissue to tissue and in different physical or pathological conditions. Figure 9 shows how the genome is linked to the phenotype through both the proteotype and the metabotype.

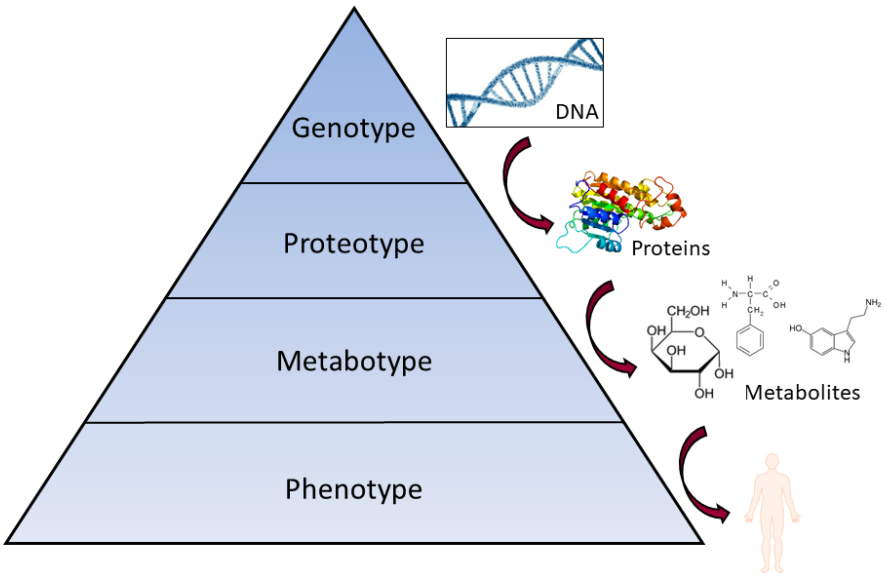


Figure 9: Scheme of the pathway from the genotype to the phenotype. Figure adapted from Liessi N., Proteomics and metabolomics for cystic fibrosis research [68].

The study of the connections between genomics variations and alterations in the phenotype can provide crucial insights into cell mechanisms, including the pathological ones.

1.2.1 Proteomics

Proteomics is the systematic characterization of the proteins expressed in biological systems, with the aim to detect their presence, their abundance, their modifications, and their interactions. The name "proteomics", coined in 1994 during a meeting in Siena [69], defines the set of proteins codified by a genome, that is the entire "PROTEins profile expressed by the genOME".

The most broadly used technology in proteomics is LC–MS (mass spectrometry coupled to liquid chromatography), which is sensitive, unbiased and compatible with large-scale analysis of thousands of proteins in complex mixtures, such as cell lysates or biofluids [70]. Protein analysis through mass spectrometry (MS) was not possible until the development of soft ionization technologies, such as electrospray ionization (ESI), for which in 2002 John B. Fenn received the Nobel Prize in Chemistry [71]. Indeed, until the 1980s, the analysis of proteins and peptides was incompatible with MS, because these "molecular elephants" (expression taken from Fenn's Nobel lecture [72]) could not be transferred into gas phase without undergoing significant degradation. Only the introduction of more appropriate ionization techniques allowed the gentle transfer into the mass spectrometer.

Even though the development of soft ionization techniques, the analysis of large and intact proteins, while growing, it is not broadly used. The standard proteomics workflow is the "*bottom-up*" approach [73], where proteins are digested into peptides before the analysis. For this purpose, sequence-specific endoproteases are used, such as trypsin, which can hydrolyze the peptide chains at the carboxyl side of arginine or lysine, except when either is bound to a C-terminal proline. The "bottom-up" approach is extremely powerful, as it enables the quantification of thousands of proteins in a complex mixture. As opposed to this approach, the "*top-down*" technique avoids the proteolytic step, analyzing intact proteins [74]. With this approach, it is possible to have information on the structure of the proteins, such as isoforms and it is generally performed on purified proteins [75].

MS-based proteomic experiments are also very useful to identify protein interactors and the post translational modifications (PTMs). Since PTMs have a crucial role in hundreds of protein functions such as cellular signaling, control of enzyme activity, protein turnover and transport, and maintaining overall cell structure [76], the power of MS in this field is manifest.

The general workflow used for bottom-up proteomics is summarized in Figure 10.

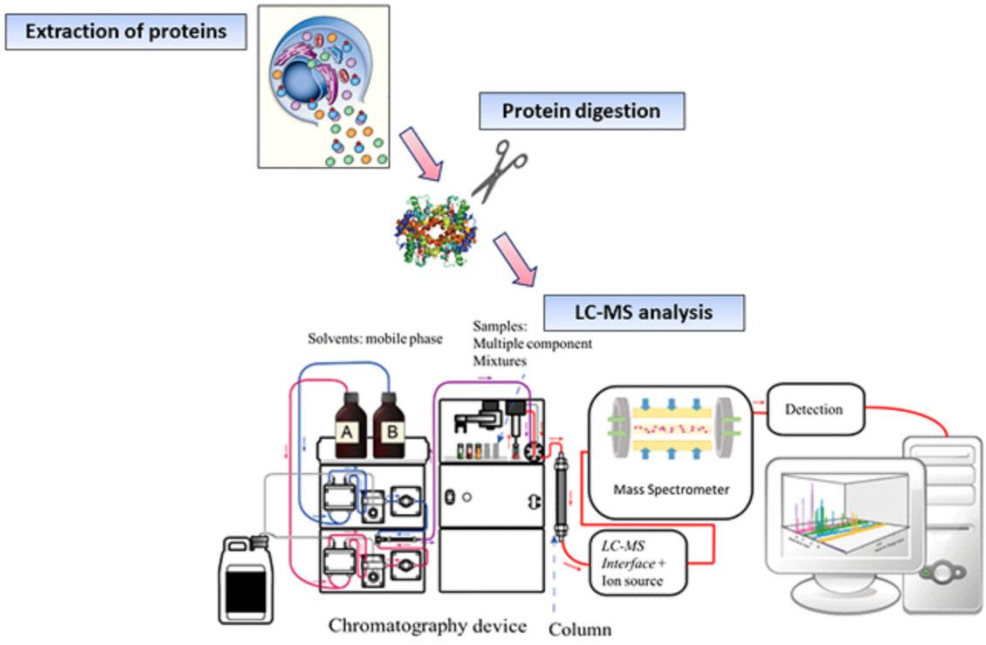


Figure 10: General workflow for LC-MS-based proteomic experiments.

Before protein analysis, it is essential to extract the proteins from the biological matrix (*e.g.*, tissues and cells) in a proper way and using the correct solvent for the lysis. Generally, the most widely used solvent for the lysis is the RIPA (Radioimmunoprecipitation assay) buffer, because of its efficiency in total cell lysis and solubilization of proteins. Indeed, this stage must ensure the efficient solubilization of as many proteins as possible, in order to achieve an unbiased and comprehensive analysis of the proteome. This step is particularly critical for membrane proteins that benefit from the use of detergents for solubilization. In order to prevent protein degradation, protease inhibitors are added to the lysis buffer. Protease activity is highly regulated in the cell: the subcellular localization of proteases is extremely important to control their activities. In fact, some proteases are confined within specific organelles (*e.g.* lysosomes) and only the proteins targeted to those organelles can be degraded. When cells are lysed for protein extraction, the membranes of the organelles are broken and proteins with different subcellular localization are released all at once, causing uncontrolled protein hydrolysis to peptides, thus resulting in decrease in the quality of MS data [21].

For the bottom-up approach, proteins are digested into peptides using specific enzymes. To allow the enzyme to reach every peptide bond on which it has to work, the 3D structures of the proteins must be unfolded. Several types of bonds are responsible for 3D-structure. For this reason, the disulfide bonds between cysteine residues are normally reduced and alkylated before protein digestion.

To increase the number of identified (and quantified) proteins, several techniques can be used. In general, the complexity of protein mixtures can be reduced by fractionation at protein level (using electrophoresis *e.g.*, a technique that separates charged proteins in an electric field) or at peptide level after protein digestion.

For proteomics, it is possible to predict peptide sequences and fragmentation spectra knowing protein sequences and enzyme cleavage patterns. Generally, thousands of spectra are acquired and, to identify thousands of proteins, specific software are needed to export fragmentation spectra and match them against *in-silico* data. The databases used for protein identification generally contain all the theoretical mass to charge ratio (*m/z*) values of peptides and their fragments, obtained by an *in-silico* digestion of all the proteins known for that specific organism of interest (*e.g. Homo sapiens*).

1.2.1.1 LOPIT-DC for spatial proteomics

In this work, proteomics was also extensively exploited to perform the Localisation of Organelle Proteins by Isotope Tagging after Differential ultraCentrifugation (LOPIT-DC) protocol. LOPIT-DC is a well-established method for the simultaneous determination of the subcellular localization of the proteins from complex biological mixtures in a single experiment. The first article describing this technique was published in 2006 [77]. Nowadays, many different versions of this protocol have been published [78]–[83].

Within the cells, proteins are spatially compartmentalized in order to precisely control cellular processes. Indeed, proteins are targeted to different subcellular compartments, where they fulfil specialized, compartment-specific functions. Three main concerns could complicate the spatial proteome profiling:

1) proteins can be present at different subcellular locations simultaneously;
2) proteins can move between different subcellular organelles and their function can vary accordingly;
3) protein localization can change during the entire life-cycle of the cells.

Therefore, the correct subcellular localization of proteins is crucial as it provides the physiological environment for their function. Protein mislocalization, together with abnormal expression in some cases, has been already associated with many diseases [84]–[86]. In some diseases, such as CF, an improper folding of a protein causes its aberrant localization. Indeed, the unfolded F508del-CFTR is retained mainly at the ER, thus impairing its functional activity. In this work, LOPIT-DC technique was optimized and applied to CF research for the first time.

Unlike microscopy-based methods, which are generally low-throughput, LOPIT-DC provides information regarding the localization of hundreds of proteins, combining biochemical cell fractionation with MS-based quantitative proteomics.

The localization of the proteins can be assigned through multivariate statistical methods based on the similarity of protein distribution profile among the fractions to the profiles of well-annotated organelle marker proteins. Indeed proteins belonging to the same organelle would show the same (or very similar) distribution profile among the fractions collected after the ultracentrifugation steps.

The workflow of LOPIT-DC technique is summarized in Figure 11.

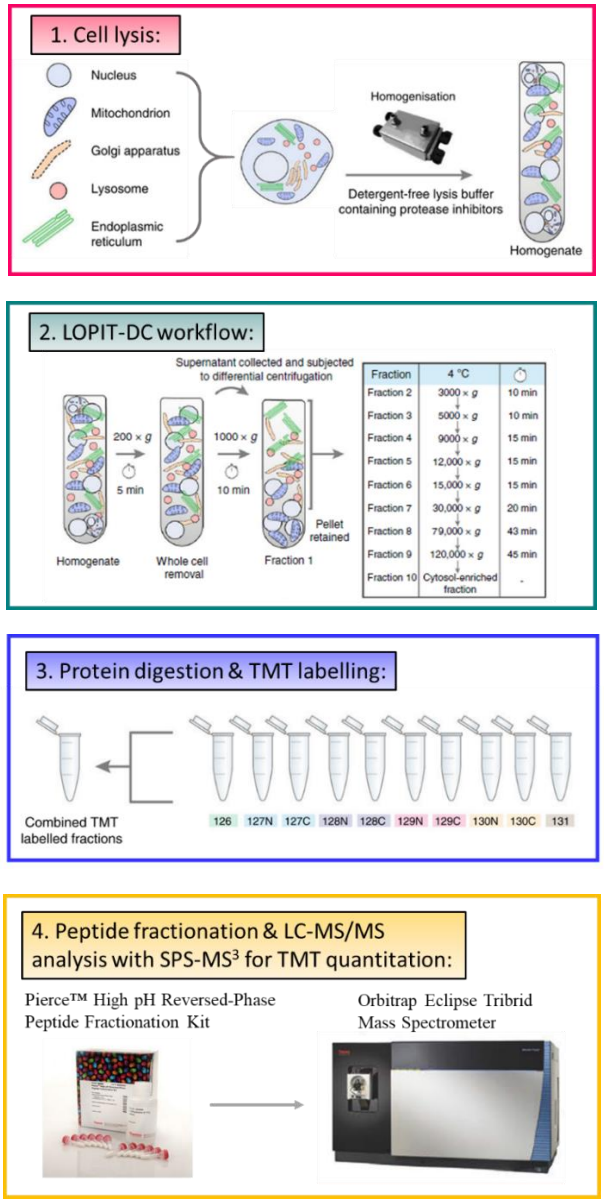


Figure 11: LOPIT-DC workflow. LOPIT is a quantitative mass spectrometry-based method used for the separation of organelles through differential ultracentrifugation steps. This workflow takes advantage of multiplex TMT-labelling (step 3) to reduce mass spectrometry analysis time and technical variability and SPS-MS³ for accurate quantification (step 4). Data analysis is performed using pRoloc (step 5) packages. Figure adapted from Geladaki A., Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics [87].

1) **Cell lysis**: the number of cells must be optimized in order to obtain at least 50 µg of proteins per fraction. The lysis is performed using the ball-bearing homogenizer and it is one of the most critical step of the entire LOPIT-DC workflow. Indeed, insufficient cell lysis can lead to reduced protein yields per fraction while excessive lysis can compromize membrane organelles making the downstream subcellular fraction impossible.
2) **LOPIT-DC workflow (ultracentrifugation steps)**: in Figure 11, the workflow used for this work is shown but speeds and times should be optimized for each type of cells.
3) **Protein digestion and TMT labelling**: protein digestion is generally followed by TMT (tandem mass tags) labelling. This is one of the quantitative proteomics techniques available in literature, perfectly suited for the multiplexed analysis of different biological samples. With this technique, the peptides are labelled with isobaric tags (their general scheme is shown in Figure 12), which bind the peptides at both N-termini and lysine side chains of peptides in a digest mixture. The labelled peptides are fragmented into the mass spectrometer and the report ions (specific for each TMT tag) are released. The intensities of each reporter ion are then used for the quantification. Each fraction deriving from the LOPIT-DC ultracentrifugation steps, is labeled with different TMT tags; in this way, the corresponding reporter ion quantification directly mirrors the abundance of that peptide in each fraction.

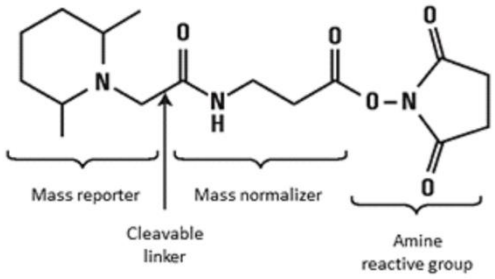


Figure 12: Each isobaric tagging reagent has the same precursor mass and is composed of an amine-reactive NHS-ester group, a spacer arm and an MS/MS reporter. The reporter ions used for quantification can be released through HDC (higher-energy collisional dissociation) or ETD (electron-transfer dissociation) fragmentation.

4) **Peptide fractionation & LC-MS/MS analysis with SPS-MS^3 for TMT quantitation**: the combined TMT-labelled peptides are then fractionated before the SPS (synchronous precursors selection)-MS^3 analysis for accurate protein quantification and spatial resolution.
5) **Data analysis (extraction of protein profile, multivariate data analysis using pRoloc package)**: data analysis for LOPIT-DC experiments is generally performed using the pRoloc, a Bioconductor R package [88]. This tool creates organelle-specific profiles using known organelle markers and matches them to the distribution profiles of the proteins with unknown localization. In the end, several algorithm can be used to assign subcellular localization of proteins, such as the support vector machine (SVM) classification [89].

1.2.2 Metabolomics

Metabolomics is the global study of all the low-molecular-weight molecules (<1000 Da) present in a biofluid, cell, tissue, organ or organism in a particular physiological or developmental state [90]. The results of metabolomic studies provide an overview of the alterations of the upstream biological processes resulting from not only the genetic background but also from environmental perturbations. Anything from small dietary changes to elevated stress can significantly alter the composition of the metabolome.

In the 1940s, Roger Williams [91], who analyzed urine and saliva to seek for metabolic patterns related to schizophrenia, performed the first "metabolomic fingerprint". In 1971 the expression "metabolic profile" was used for the first time [92], when advancements in technology allowed the shifting from qualitative to quantitative measures of metabolites.

As opposed to proteomics, where the molecules to be analyzed are essential similar from a chemical point of view, metabolomics aims at tracking simultaneously multiple small molecules types, such as lipids, carbohydrates, amino acids and other products of cellular metabolism.

Due to the chemical diversity of the metabolites, LC-MS is not the only analytical technique used for metabolomics (as it happens for proteomics). In fact, metabolomics is also frequently performed by both GC-MS (gas chromatography coupled to mass spectrometry) and NMR (nuclear magnetic resonance) spectroscopy. The choice of the technique for the analysis depends on the experimental objective and sample time. Despite being one of the oldest MS-based techniques, GC-MS still offers unmatched performances for the identification (and quantification) of small and volatile molecules. Nevertheless, its use is limited because it requires the chemical derivatization of all the non-volatile biological molecules [93]. Even though much less sensitive than MS, NMR is the state-of-the-art technology to analyze polar compounds. Moreover, with NMR it is possible to perform absolute quantification of the metabolites (not possible with MS).

In contrast with other studies (*e.g.*, genomics) but in common with proteomics, there are no methods for the amplification of the metabolites, thus making the sensitivity a major issue in metabolomics [90].

Before any analysis, the metabolism in the sample matrix needs to be stopped to prevent metabolites degradation. Since their turnover can occur within seconds [94], strategies must be put in place to avoid significant alterations during the sampling. The block of the metabolism reactions can be achieved by freezing and/or enzyme denaturation, following one of the many protocols available in literature [95], [96]. Generally, for cells, adding cold organic solvents can block all the enzymatic activities, while for tissues it is better to freeze first and then extract. The general workflow for metabolomic analysis is summarized in Figure 13.

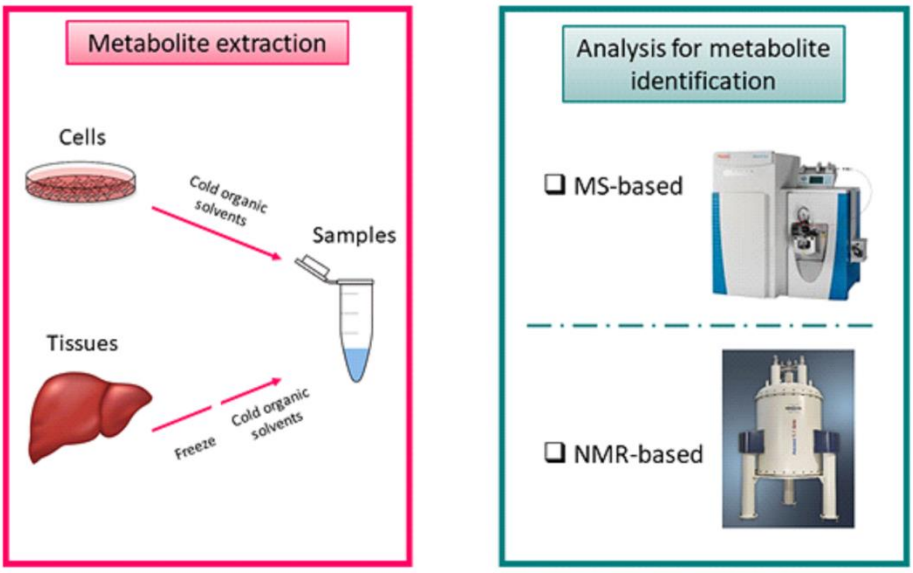


Figure 13: Scheme of the general workflow for metabolomic experiments.

Once extracted, the metabolites must be analyzed using the more appropriate instrument able to detect as many metabolites as possible. Metabolites are much more challenging to annotate than proteins because of the lack of a genetic template. For this reason, validation of retention times and fragmentation spectra is often required to have confident metabolite identification.

Untargeted approaches, generally used to generate hypotheses, focuses on acquiring and analyzing as many metabolites as possible to detect metabolic changes. With this approach, data can be used for relative quantification thus providing a trend of the distribution of both the unknown and known metabolites across the samples. The results of untargeted analyses can generate hypotheses that should be validated with targeted approaches, whose aim is to identify and quantify (absolute quantification) a limited number of known metabolites. Figure 14 shows the different goals of targeted and untargeted analyses.

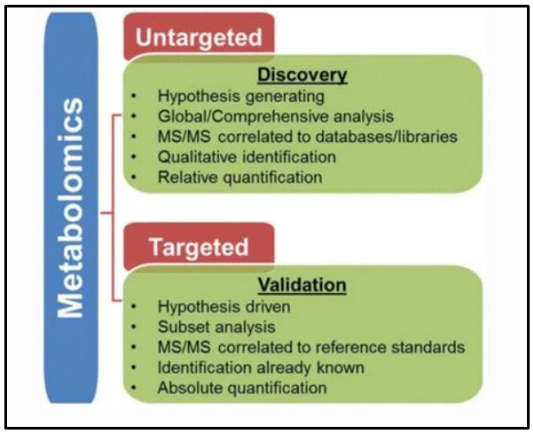


Figure 14: Comparison of untargeted and targeted metabolomics studies. The discovery-based untargeted metabolomic studies aim at global detection and relative quantitation of small molecules. Targeted metabolomics is generally used to validate results obtained in untargeted approaches. The aim of targeted metabolomic studies is to focus on measuring well-defined groups of

metabolites with opportunities for absolute quantitation. Picture taken from Schrimpe-Rutledge A. C.: Untargeted Metabolomics Strategies—Challenges and Emerging Directions [97].

The general workflow for untargeted metabolomics data analysis is summarized in Figure 15.

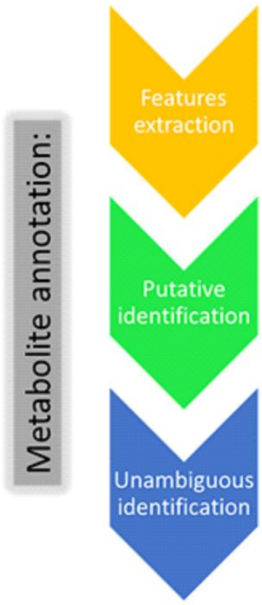


Figure 15: Scheme of the steps for metabolite annotation.

The first step is generally the ***features extraction***: untargeted LC-MS metabolomic studies can detect thousands of signals, commonly called features [98]. Peak-picking algorithms [99]–[101] detect those signals that can be associated with metabolites. Since, molecules are frequently detected in electrospray mass spectrometry in the form of different adduct (*e.g.*, $[M+Na]^+$, $[M+NH_4]^+$, $[M+HCOO]^-$), besides protonated or deprotonated species, these algorithms also help in the deconvolution of the observed features.

Only a small portion of the detected features can be readily identified. The real bottleneck of the data analysis in metabolomics is the ***compound identification*** step: the *m/z* values are searched against metabolite libraries [102]–[104] to obtain a list of putative identifications. The metabolites having molecular weights within a specified tolerance range to the query *m/z* value are retrieved from the databases as putative identifications.

Finally, whenever possible, the MS/MS spectra and the retention times of the putative metabolites are compared with the standard compound analyzed under identical conditions for the ***unambiguous identification***, according to the Metabolomics Standard Initiative (MSI) guidelines [105].

1.3 Omics in CF research

Both proteomics and metabolomics were applied to cystic fibrosis research for many applications: from the identification of CFTR interactors to biomarker discovery [68]. The final aim of all these studies is to elucidate the phenotypes resulting from CFTR mutations at both metabolomic and proteomic levels.

Proteomics was successfully applied to study CFTR interactome to have a comprehensive overview of how CFTR works inside the cells. The first interactomic study was performed in 2006 by Wang *et al*. [106]. The authors exploited their newly-developed technique - the multidimensional protein identification technology (MudPIT) - to study the wt-CFTR interactome in different cell lines (both lung and intestine cell lines). Moreover, they applied the MudPIT workflow to study the different proteins present in the wt-CFTR interactome in comparison to the F508del-CFTR interactome. In particular, they focused on finding differences in proteins involved in ER machinery transport. This experiment led to the identification of the Hsp90 cochaperone aha1 as a specific F508del-CFTR interactor.

Other interactomic experiments were conducted to study the different interactome of wt-CFTR (expressed in HBE14o-) and F508del-CFTR (present in CFBE41o-) in 2015 [107]; they also studied how the interactome changed after F508del-CFTR rescue maneuvers. This study showed that wt- and F508del-CFTR share 85% of their interactomes. Among the proteins that specifically interact with F508del-CFTR, chaperones involved in ER quality control and in protein degradation were identified. Interestingly, this study identified also a set of proteins involved in the endoplasmic-reticulum-associated degradation (ERAD) but never associated to CF (such as the proteins AUP1, SEL1L and FAF2).

In 2018, MS-based proteomics coupled with bioinformatics tools were applied to study how F508del-CFTR is retained in the ER by the ERQC (ER quality control) machinery [108]. In particular, the interactome of F508del-CFTR at the AFT motif was investigated. As already described in Paragraph 1.1.3, the AFT motif is used by the cells as a signal to retain unfolded proteins in ER (the exposure of this motif is regulated by the folding status of the protein). Among the proteins that specifically interact with F508del-CFTR at the AFT motif, proteins involved in the vesicle-mediated transport were identified as well as proteins involved in the folding mechanism [108].

The last interactomic experiment was performed in 2019 [109], leading to the identification of Calpain-1 protein as F508del-CFTR interactor at the plasma membrane. In this study, engineered immortalized bronchial cell line (mChery-FLAG-CFTR CFBE cells) were used to develop a particular immunoprecipitation-based protocol containing a cross-linking agent. With this new immunoprecipitation method, the authors were able to select the CFTR interactome at the plasma membrane. Three proteins were selected as relevant interactors belonging to the NHERF1-ezrin complexes: 14-3-3 zeta, calpain-1 and importin 5. The particular role of calpain-1 in reducing the F508del-CFTR stability at the plasma membrane was further investigated: they demonstrated that a down-regulation of calpain-1 triggers a functional rescue of F508del-CFTR in airway epithelium.

Proteomics was also applied to study PTMs and their influence on CFTR activity. The first studies on PTMs demonstrated that misfolding of the mutant CFTR impaired the PTMs-regulated activities of the channel [12] and only in 2019 the PTMs were described as key regulators of CFTR biogenesis and trafficking [110]. In particular, they studied how some PTMs can regulate the turnover of CFTR and how different is the PTMs landscape of CFTR with different CF-causing mutations. Moreover, they investigated if differences in PTMs can have a role in CFTR rescue. Their findings showed that the amount of phosphorylation in

G551D and R117H mutants (having a defective ion channel activity with no effects on biogenesis) was comparable to the amount of phosphorylation in wt-CFTR. On the contrary, the phosphorylation at Thr421 to Ser427 was drastically reduced for N1303K- and F508del-CFTR (both having a defective maturation), suggesting that these phosphorylation sites are crucial for CFTR biogenesis.

Only few studies were performed to detect global changes in protein expression profiles in CF bronchial tissue.

Two studies were performed exploiting the 2DGE/MS-based technology, which has a bias toward the most abundant and soluble proteins:

I. In 2006, protein expression of bronchial biopsies derived from 9 CF patients were compared to those of 8 control subjects [111]; this study revealed that proteins involved in inflammation, infection and cellular stress response were differentially expressed.
II. A more recent study, in 2018 [112], showed that 15 proteins were significantly dysregulated by the F508del mutation. The proteome of CFBE41o- cells, expressing F508del-CFTR, was compared to that of 16HBE14o- cells, expressing wt-CFTR.

The first shotgun, label-free quantification of protein from total lysates of cells was performed in 2014 [113]. The proteome of CFBE41o- cells was compared to that of 16HBE14o- cells: among the thousands of proteins detected - while with 2DGE/MS methods hundreds of proteins are detectable – 349 proteins were found to be significantly dysregulated by the F508del mutation.

The first (and only) quantitative proteomic study on primary bronchial epithelium cells was performed in this thesis work, as described in (Paragraph 4.2).

Even though these studies provided insights into some CF-related processes, very few papers report data from a large-scale systematic investigation. The same happens for metabolomic studies, which were conducted mainly for biomarker discovery rather than to investigate basic CF pathophysiology at the cell level.

The first untargeted metabolomic study was performed in 2010 [114] to pinpoint the epithelial dysfunction caused by CF. They used primary airway epithelium cells to compare the metabolome of CF patients to non-CF subjects, revealing some metabolic alterations:

- Nucleotide metabolism, with a decrease in purine biosynthesis, essential for the control of ASL (airway surface liquid) volume.
- Increase in the catabolism of Tryptophan, causing accumulation of molecules associated with oxidative stress.
- Reduction of glutathione biosynthesis, a key regulator of the oxidative status of the cells.
- Decrease in osmolytes, such as sorbitol and glycerophosphorylcoline, which have a role in the regulation of cell volume.
- Low levels of glucose metabolism, possible cause of an increase in cell sensitivity to oxidative stress.

Other contributions to the understanding of the molecular mechanisms of CF disease were provided by Bear's group who studied the role of cholesterol and sphingolipids on CFTR stabilization and regulation [115].

Further metabolomic studies were performed comparing the serum of 31 CF children to the serum of 31 non-CF children [116]. 92 metabolites were found to be dysregulated by CF, revealing some novel pathways dysregulated in CF (such as cellular energy production through the β-oxidation of fatty acids) together with already known altered pathways (*e.g.* increased oxidative stress).

The metabolome of serum was also studied to monitor the efficacy of Orkambi (the combination of Lumacaftor and Ivacaftor) treatment [117]. Despite the limited number of samples (only 20 patients were monitored after a six-month therapy), metabolites involved in lipid and aminoacid metabolisms were found to be altered by the treatment.

Together with proteomics and metabolomics, transcriptomic studies were also performed to better understand how dysfunctional CFTR impacts on transcriptional events in CF. Both microarray and RNA-sequencing techniques were exploited in many CF studies, with different aims:

- to investigate transcriptional changes induced by bacterial infections [118], [119].
- to compare CF patients to non-CF subjects [120].
- to study changes associated to pharmacological treatment, such as Azitromycin in 2009 [121] or CF correctors/potentiators (in 2019 only with Ivacaftor to predict clinical responsiveness [122], in 2020 the treatment with Orkambi was investigated [123]; treatment for exacerbations: in 2018 [124])
- to study changes after external *stimuli*: oxidative stress in 2014 [125], digitoxin in 2017 [126], genomic variations in 2018 [127].

All the cited studies were performed using bronchial epithelial cells, as they are the most affected by the impaired function of mutant CFTR in the lungs. Most of these studies detected as altered several pathways related to inflammation, such as NF-KB signaling [119], [120]. Interestingly, transcriptomics investigation of the effect of Azytrocyin showed alteration in the lipid metabolism [121].

1.4 LIQUID CHROMATOGRAPHY - MASS SPECTROMETRY (LC-MS) for omics

Mass spectrometry is a powerful tool to detect and quantify biomolecules in a complex mixture because of its sensitivity and relatively high-throughput. MS is generally coupled to liquid chromatography (LC), that is a technique commonly used to separate the analytes (peptides for bottom-up proteomics and metabolites for metabolomics) dissolved in a solvent called "**mobile phase**" or eluent. With the use of pumps and an injection system, the eluents bring the analytes to the detection system passing through a "**stationary phase**". The chemical proprieties of each analyte directly results in its partition between the stationary and the mobile phases.

Based on the different chemistries of the interaction of analytes with the stationary and the mobile phases, several types of chromatographies are available for the coupling with MS. The most widely used chromatography for "omics" applications is the "reverse phase liquid chromatography" (RP-LC), which is able to separate the molecules mostly based on their hydrophobicity, generally indicated by the LogP (the logarithm of the partition coefficient defined as the ratio of the concentrations of the compound in a mixture of two immiscible solvents). In RP-LC, the analytes elute mainly based on their increasing LogP value, with the eluent phase composition changing over time (gradient separations).

After sample loading into the chromatographic column, the analytes are separated and eluted through a gradient of elution strength in the mobile phase, usually achieved with the use of acetonitrile or methanol. The separation of the molecules contained in complex mixtures occurs because each of the molecules is eluted at a different time (corresponding to a different concentration of mobile phase), depending on its chemical composition. The time of the elution of a molecule is called retention time.

Once eluted, the analytes enter the mass spectrometer, which is usually made of three main parts:

1) **Ion source**, where the analytes are transferred as ions from aqueous to the gas phase. The electrospray ion source (the most widely used for omics applications) nebulizes the analytes in small droplets which are ionized by a potential difference able to confer a charge to the droplets [128].
   Electrospray ionization (ESI) allows both the ionization and the transfer into the gas phase of molecules in solution, as shown in Figure 16. In LC-MS, the electrospray is applied at the end of the chromatographic column, by applying a high voltage to the eluted mobile phase, causing the production of small, charged droplets. The evaporation of the solvent from the nebulized droplets results in the so-called "coulombic explosion", with ions of the same charge ejected in the gas phase as "naked" ions [129]. Only charged molecules can be analyzed by the mass spectrometer, while the instrument does not detect neutral molecules, which are generally prevented from entering the mass analyzer.

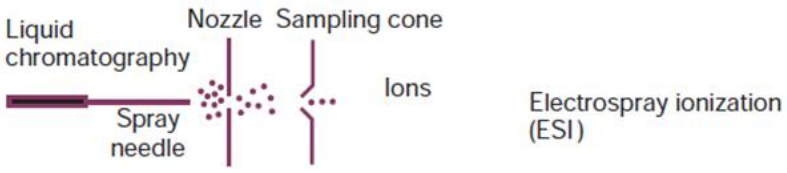


Figure 16: Schematic representation of an ESI ion source. Picture adapted from Aebersold R. Mass spectrometry-based proteomics [128].

2) **Mass analyzer**, which measures the mass-to-charge ratio (*m/z*) of the ions. The ability to separate molecules based on different *m/z* was first described in 1912 by Thomson, the winner of the 1906 Nobel Prize in physics for his studies on the conduction of electricity by gases [130]. There are different types of mass analyzers. Each of them is characterized by distinctive **resolution** and **mass accuracy**.
The instruments used for this thesis belong to the so-called "*hybrid*" mass spectrometers class, to obtain both high mass accuracy and good sensitivity during the detection of the ions.
In fact, they have both the quadrupole and the time of flight (TOF) mass analyzers, as illustrated in Figure 17. The first quadrupole (Q_1) acts as a filter to select a particular *m/z* value. The second quadrupole (q_2) works as a collision chamber, where the ions are fragmented. Once arrived in the TOF analyzer the charged molecules (either intact or fragmented) are accelerated to high kinetic energy and are separated along a flight tube, as a result of their different velocities, which depend on their mass (with lower *m/z* ions flying faster).

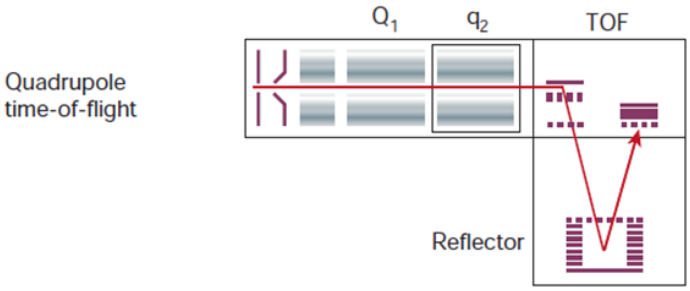


Figure 17: Quadrupole-TOF analyzer scheme. Picture adapted from Aebersold R. Mass spectrometry-based proteomics [128].

In this work, another mass analyzer was used: the Orbitrap, shown in Figure 18. This relatively new (since 2006, [131]) mass analyzer is able to monitor the oscillations of the ions that orbit around a central spindle-like electrode with electrostatic attraction balanced by centripetal force. The frequency of the oscillations of the ions along the axis of the electrode is characteristic of the ion *m/z* value.

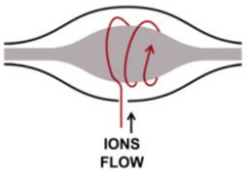


Figure 18: Ions movement in the orbital analyzer.

3) **Detector**: a device able to convert the arrival of the ions into a detectable and measurable ion current.

The result of the combined work of these three devices is a histogram of intensities at different *m/z* values, called "mass spectrum".

Most mass spectrometers have also the ability to fragment the molecule and to acquire the corresponding fragmentation spectrum (MS/MS or tandem MS or MS^2). The fragmentation peaks produced by MS/MS experiments are used for the identification of the biomolecule.

Different types of instruments, having different ion sources and mass analyzers, are currently used in the MS-based omics field.

1.4.1 qTOF Mass spectrometers

For all the expression proteomic experiments reported in this work, a SCIEX TripleTOF5600+ was used. The diagram of this instrument is illustrated in Figure 19.

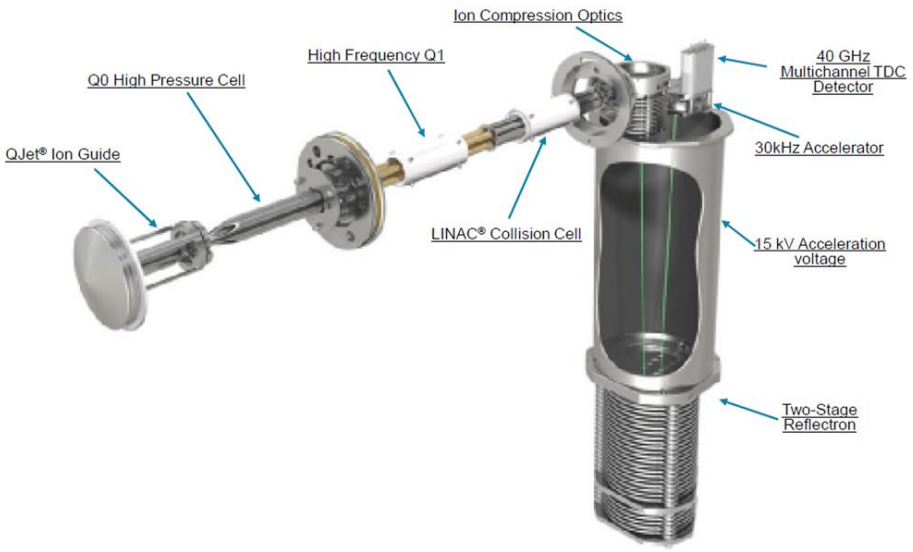


Figure 19: Scheme of the components of the TripleTOF5600+ instrument.

This instrument has a series of quadrupoles aiming at filtering and focusing the ions and at removing the neutral molecules. This increases the sensitivity of the instrument and the signal-to-noise ratio. The first quadrupole is the QJet ion guide that focuses the ions before their entrance into the second quadrupole of the series, the Q0. In the Q0 region, the ions are focused again and only the Q1 quadrupole filters the ions before they enter the Q2 collision cell. In this cell, collisions with gas molecules cause the breaking of molecular bonds, through the conversion of the ion kinetic energy into vibrational energies of the chemical bonds. Finally, the ions arrive at the TOF analyzer where they are separated in time based on their *m/z* ratio.

This instrument has two operating modes: the Data-dependent acquisition (DDA) and Data-independent acquisition (DIA).

In the DDA approach, as the compounds enter the mass spectrometer, the instrument quickly scans them and chooses a subset of signals for the fragmentation. The choice of the instrument relies on some criteria previously set up in the method, which is usually based on the ions' abundance or, particularly in proteomics, their charge state. In general, with this method it is possible to select a number of the most abundant ions for the fragmentation. The MS/MS spectra derived by the fragmentation are used for further identification.

Unlike data-dependent acquisition (DDA) methods, where only a subset of a fixed number of precursor ions (the most abundant ones) are fragmented, Sequential Window Acquisition of all Theoretical fragmentation spectra (SWATH) MS (mass-spectrometry) is a data-independent acquisition (DIA) where all the ionized peptides are fragmented and can be identified [132]. In fact, SWATH allows the fragmentation of (in principle) every peptide present in a complex mixture within a broad and pre-defined mass range.

Sequential Window Acquisition of all Theoretical fragmentation spectra [133] is a label-free quantification method that allows a complete and permanent recording of all the fragment ions of the detectable peptide precursors, thus leading to the quantification of several thousands of proteins in a biological sample.

The major advantage of SWATH MS is that it combines the high reproducibility and sensitivity of targeted methods with the huge proteome depth typically obtained with DDA MS. Thanks to its accuracy and reproducibility, SWATH is suited for projects with a large number of samples, which require accurate and reproducible quantification.

To query the SWATH data, where theoretically all the detectable peptides in the sample can be identified, a library containing previously acquired data is used. The libraries are built in DDA mode and report information regarding the chromatographic and mass spectrometric behavior of all the detected peptides in the biological matrix.

Nevertheless, the generation of a project-specific ion library makes SWATH less attractive for studies with a low number of samples. The other drawback of this technique is that only proteins which assays are reported in the IL can be quantified [134]. The quality and coverage of such a reference library are therefore exceedingly important.

In the context of SWATH proteomics, an “assay” is a text string that reports the observed molecular information for a given peptide (accurate mass, retention time and, most importantly, its corresponding MS/MS fragments with their intensities). The assays reported in the library are used by the data analysis software to univocally link a given set of peptides to their corresponding originating protein, thus allowing its simultaneous identification and quantification.

For the metabolomic analysis of this work two different instruments were used: SynaptG2 for untargeted metabolomics and a triple quadrupole for targeted analysis.

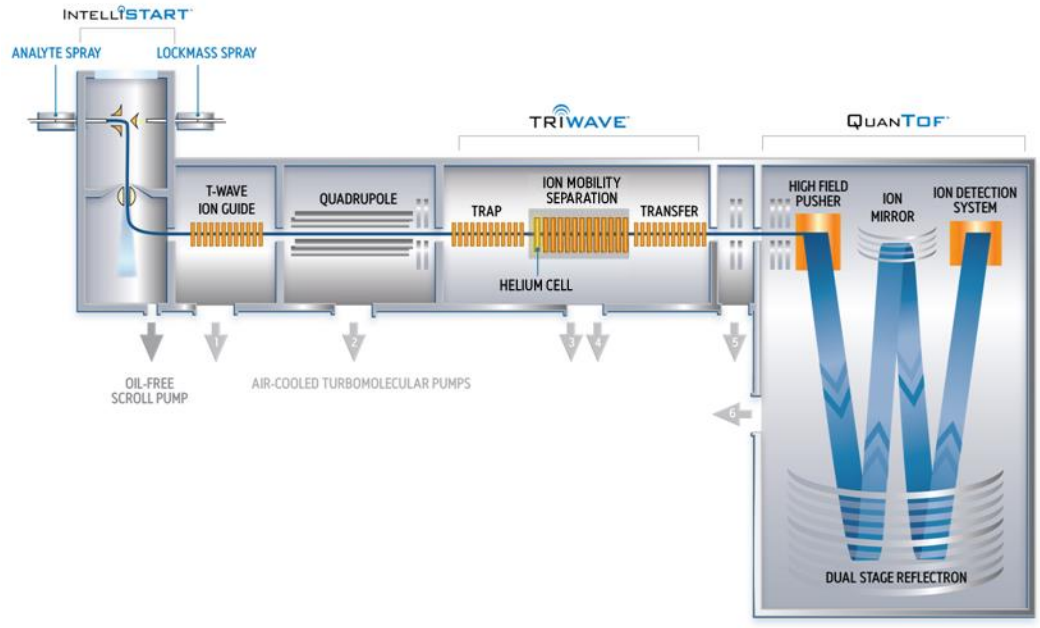


Figure 20: Scheme of the components is Synapt G2 QTOF high-resolution mass spectrometer. This instrument was used in this work for lipidomics untargeted analysis.

As shown in Figure 20, the QTOF Synapt G2 is a hybrid mass spectrometer, where the quadrupole selects and fragments the ion before they arrive at the TOF analyzer. The ion guide before the quadrupole has the capability of focusing the ions and removing the neutral contaminants thus increasing the sensitivity of the instrument. For untargeted analysis, this instrument works acquiring in MS^E mode [135], where full MS^1 scan is integrated with MS/MS fragmentation for all the precursor ions.

For the targeted analysis present in this work a quadrupole was used. The general scheme of a quadrupole is shown in Figure 21.

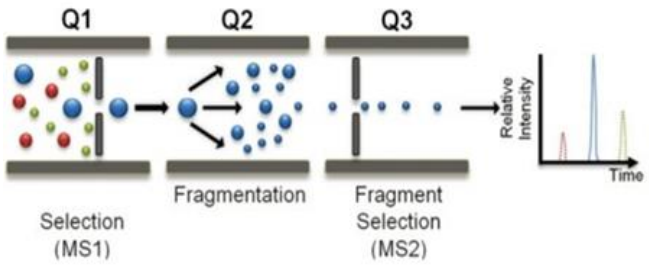


Figure 21: General scheme of a triple quadrupole mass analyzer.

For targeted analysis, the selected molecules to be analyzed are filtered by the first quadrupole. The second quadrupole acts as a collision cell for the fragmentation of the selected molecules through collision-induced dissociation (CID). In the end, the third and last quadrupole allows the detection of only distinct fragment ions deriving from certain precursor ions. This configuration is called multiple reaction monitoring (MRM) [136].

1.4.2 Eclipse for LOPIT-DC experiments

For the LOPIT-DC experiment, the Eclipse Tribrid Mass Spectrometer was used. The major components of this instrument are summarized in Figure 22.

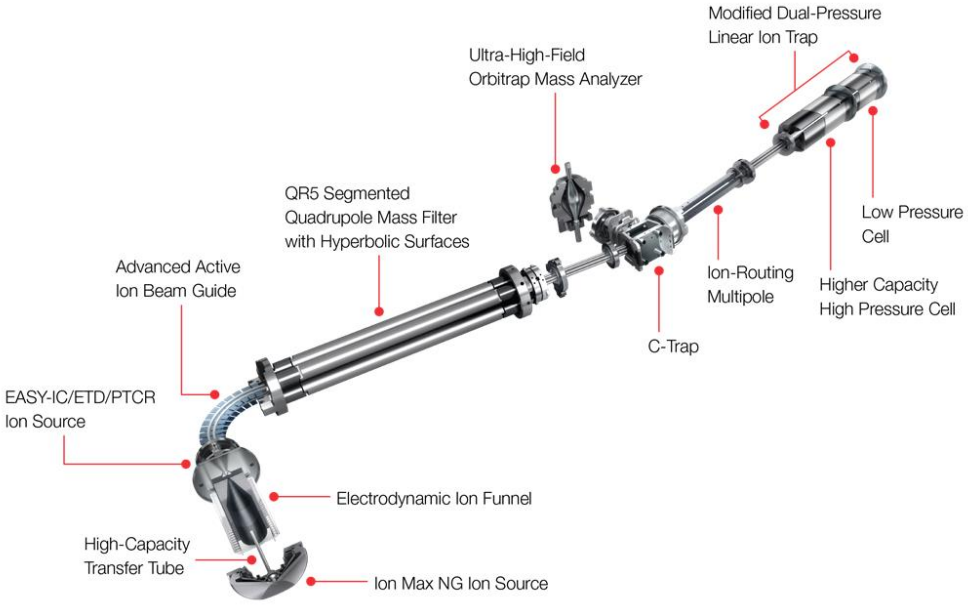


Figure 22: Eclipse ion path.

With this instrument, three different dissociation techniques (CID, collision-induced dissociation; HCD, higher-energy collisional dissociation; and ETD, electron-transfer dissociation) can be performed at any fragmentation stage, followed by the acquisition by either the ion trap or the Orbitrap mass analyzer. Moreover, the TMT quantification can be performed following the real-time search SPS (synchronous precursors selection)-MS^3 workflow [137]. This approach relies on searching MS/MS spectra in real-time during the acquisition thus performing SPS-MS^3 only for identified spectra eliminating the MS^3 acquisition on unidentified peptides.

1.5 Multivariate data analysis

Modern omics experiments generate large amount of data. Therefore, proper data analysis and proper statistical methods are needed to obtain trustworthy results. Because of the big amount of data, bioinformatics tools are essential to process this kind of data [138].

The general workflow for data analysis aims at integrating omics data with annotational databases, such as Gene Ontology (GO) [139] or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [140] to yield functional insights into the data.

The output of quantitative omics experiment is a data matrix where for each sample thousands of features (variables) are quantified. To extract information from data with multiple variables, looking at all the variables simultaneously, multivariate data analysis without losing any information is needed. This tool enables also the visualization of the matrix generated by omics.

When a biological system is perturbed – *e.g.*, by genetic editing or pharmacological treatment – several changes might be observable, in principle. Advanced data analysis methods are used to discriminate which changes are associated to the perturbation we applied. For this reason, further biological investigations are always required to validate the results of omics experiments.

Several statistical algorithms and software have been developed in recent years to deal with the complex dataset generated by quantitative comparisons between two or more biological conditions. For a sophisticated data visualization and statistics, several software were used in this work such as Metaboanalyst [141], Markerview [142], and the open source Bioconductor within the R statistical platform [143], which requires programming experience.

For a more accurate quantification, data must be pre-processed through normalization and scaling.

The ***normalization*** aims at reducing technical sample-sample variability without affecting biological variability. There are several ways to perform normalization and the right choice depends on the experimental design; the main normalization techniques are:

- Using internal standards (IS): used when a specific feature is used to calculate and apply new normalization factors across all the samples of the dataset. For each sample, the normalization factor is calculated as the ratio between the average response of the IS for all the samples and the response in the sample. The final normalization factor is the average of all the normalization factors.
- Using total area sums: each sample is normalized so that the normalized sample has the same area sum calculated using all peaks. This kind of normalization is useful when most of peaks in the samples are expected to have similar intensities in all the samples.
- Using the most likely ratio (MLR) [144]: used especially when a large number of features (and thus variables) are detected – the ratio between features are used to compute a ratio histogram which is aligned across samples.

Finally, the ***scaling*** aims at reducing substantially different ranges that variables can have across the sample. If the original data is not scaled, the variables with the largest response will tend to dominate the PCA analysis. Three main scaling approaches can be used:

- Mean centering: after this scaling, each variable retains its original range, but its average value is centered at zero.
- Autoscaling: the data is first mean centered and then divided by the standard deviation for the variable. The final effect is that each scaled variable has a mean of zero.
- Pareto scaling: this is a compromise between mean and autoscaling and it is often used in data analysis for omics. The data is first mean centered and then divided by the square root of the standard deviation for the variable.

After the pre-processing of the data, an essential tool to visualize and interpret data is the Principal Component Analysis (PCA). PCA is an unsupervised multivariate statistical analysis, aiming at the dimensionality reduction of the original multi-dimensional matrix data into a low number of dimensions, called principal components (PC). This reduction allows the visualization of complex dataset. PCA allows the recognition of trends across groups of sample within a dataset. The first principal component (PC1) accounts for the highest difference between the samples, PC2 (orthogonal to PC1) accounts for the second highest difference and so on (with PC3, PC4…). The result of this analysis is a PCA score plot where each sample is represented by a dot in a two-dimension space (PC1 x-axis and PC2 y-axis).

A supervised version of PCA (the partial least squares discriminant analysis, PLS-DA) can also be used to identify the variables that discriminate the groups with a prior knowledge of sample groups. This supervised technique uses class information to improve sample preparation: PLS-DA is a one-step procedure that combines dimensionality reduction and discriminant analysis into one algorithm.

2 Rationale and aim of the project

The development of new effective therapies for CF increased the life expectancy of CF patients over the past three decades: whereas in 1980s the median survival for CF patients was only 16 years, nowadays life expectancy for people with CF can exceed 40 years [145].

To improve the activity of mutant CFTR, two strategies are essential: increasing the number of the channel available in the plasma membrane and its opening time. The new CFTR modulators work at both levels separately: the potentiators aim at increasing the open channel probability, while the correctors improve the folding and, thus, the trafficking of the mutant CFTR at the plasma membrane.

As discussed in the introduction, the combination of molecules exploiting different mechanisms of action is needed to achieve a therapeutically relevant rescue of CFTR activity. Indeed, it is not sufficient that an adequate amount of mutant CFTR reaches the PM, it should also work properly when there.

The new CFTR modulators – both the correctors and the potentiators – have not the same clinical outcomes for all the mutations. For example, both the Symdeko and the Orkambi can be used for the treatment of CF patients homozygous for F508del-CFTR mutation [146]. However, none of these drugs is sufficiently effective in CF patients carrying a single F508del-CFTR allele and a second CFTR mutation that does not respond to the current CFTR modulators therapy [147].

Despite all the recent progress in CF therapy, there is still work to be done in order to help all the people with CF benefit from the new modulators. For this reason, expanding the portfolio of molecules for CF treatment could be extremely useful to help CF patients in living longer and healthier lives.

To expand the portfolio of molecules available for CF patients and exploiting different molecular mechanisms of action, finding new targets for CF pharmacological treatment would be extremely useful.

For this reason, the aim of this work is to use omics experiments to find new potential targets for CF pharmacological treatment. For this purpose, omics were used to provide an insight into the changes in the cell molecular space associated with successful CFTR rescue maneuvers. The rescue was achieved by the use of drugs or by selectively silencing the expression of other proteins, able to improve the trafficking of CFTR itself to the plasma membrane. The general rationale for the project follows.

If, by using omics, we are able to monitor with a sufficient analytical depth the chemical space composition of proteins and metabolites within the cell and we are able to accurately monitor their alterations following CFTR rescue, then we will gain a crucial understanding of how, at the deepest biochemical level, the cell itself is reacting to the rescue. For example, if, with proteomics, we identify a set of proteins that are downregulated following CFTR rescue, we can then hypothesize that this association works in both ways. Perhaps, by inhibiting the action of those same proteins, we will favor the rescue itself.

Two different CFTR rescue strategies were used:

1) We first used siRNA to selectively silence a set of proteins, named now “Primary targets”, whose inactivation is known to be associated with a positive CFTR rescue [148]. The resulting proteins and pathways significantly altered by the genetic rescue were investigated to find new potential targets. Some new target candidates were selected and compounds aiming at their inactivation were first tested on CFBE41o- cells and then on the bronchial epithelium (BE) primary cells to validate the findings. The BE primary cells were also analyzed to detect proteins (and pathways) altered by the disease.

2) We then used a pharmacological approach to obtain the rescue: we incubated CFBE41o- cells with a set of drugs and drugs combinations, including Kaftrio, which are currently used by CF patients [149]. We then used MS to monitor the corresponding alterations on the cell lipidome in response to these stimuli.

Finally, we applied a very advanced omics investigation that was never used before for CF research: to monitor how a pharmacological rescue of CFTR alters the subcellular localization of thousands of proteins within CFBE41o- cells we thus used the LOPIT-DC workflow.

A scheme of all the experiments performed in this work is shown in Figure 23.

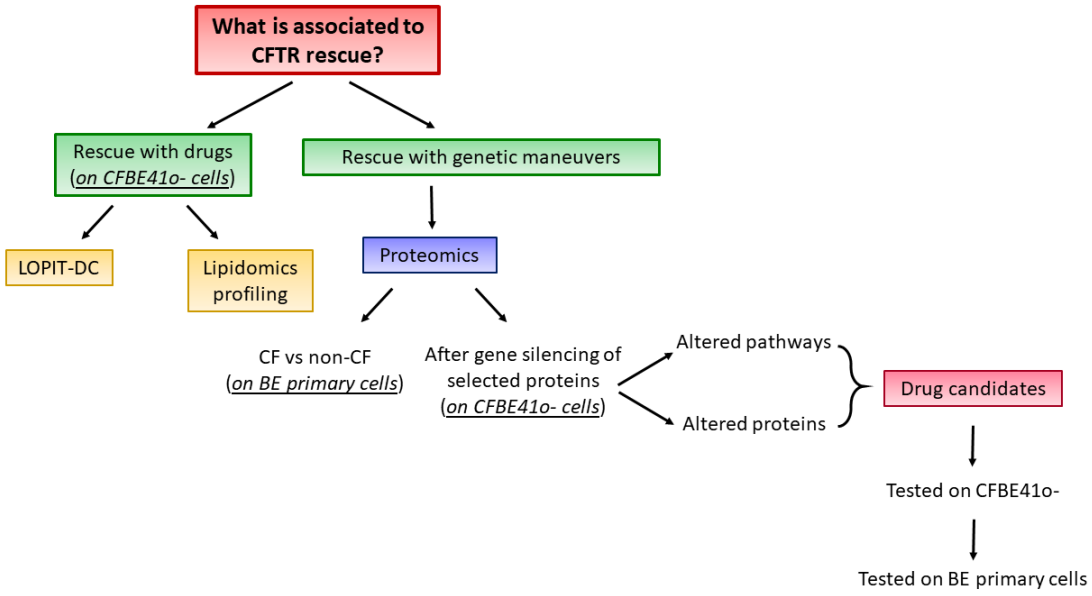


Figure 23: Summary of all the experiments performed in this thesis work.

3 Materials and Methods

3.1 Creation of the ion library for CF research

As described above (Paragraph 1.4.1) SWATH proteomics relies on ion libraries for protein quantification. The Pan Human ion library, created by Rosenberger and colleagues [150], currently represents the state-of-the-art tool for this kind of analysis on human samples. Despite its outstanding quality, the Pan Human ion library lacks assays for several proteins crucial for the CF biology, including CFTR. To gain a better understanding of CF disease, a specific ion library was built [151].

For this analysis, the most popular *in-vitro* model to study CF was used, *e.g.*, the CFBE41o- cells. Moreover, to be more specific, the BE primary cells were also used. The general workflow to create the CF-specific ion library is summarized in Figure 24.

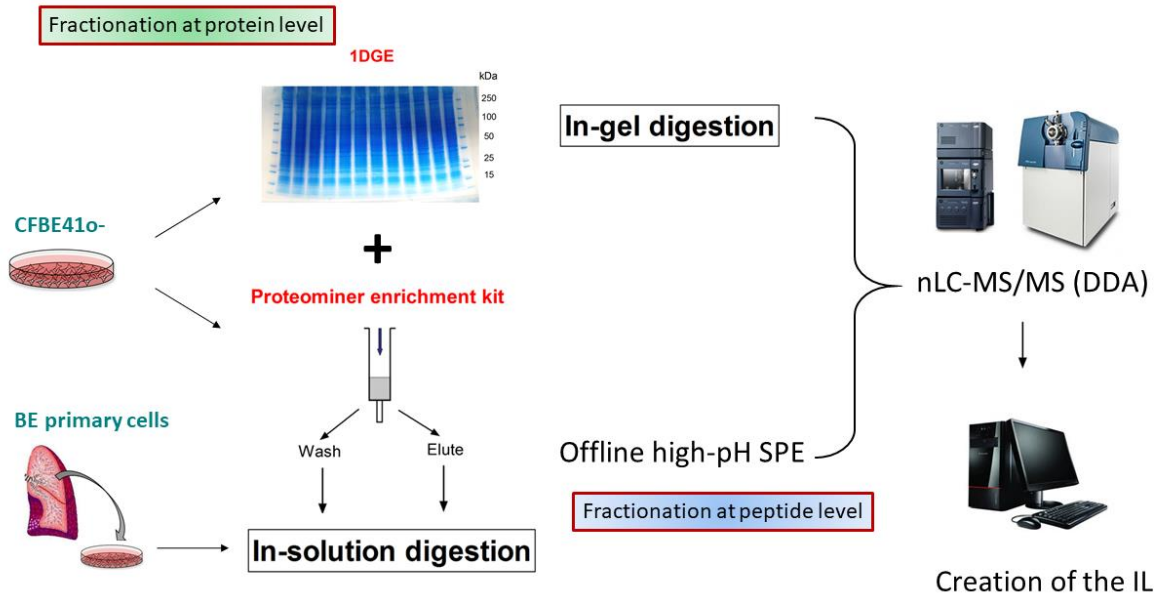


Figure 24: Schematic workflow for the creation of the ion library.

With the aim of increasing the total number of detected proteins, two kinds of fractionations were used, both at protein and peptide levels. This double fractionation strategy was possible only in CFBE41o- cells, as they grow quickly and continuously and it is possible to extract large amounts of proteins. For the BE primary cells, only one fractionation method was used instead.

3.1.1 CFBE41o- cell cultures

In order to minimize possible differences due to clonal variability, three different CFBE41o- cell populations were pooled: CFBE41o- stably over-expressing the F508del CFTR mutation, or the wt-CFTR allele, and isogenic CFTR null cells (CFBE41o-, null). CFBE41o- cells were cultured, transfected and grown to confluence on 60-mm diameter dishes and then lysed in RIPA buffer added with complete protease inhibitor (Roche). Cell lysates were subjected to centrifugation at 15000xg at 4°C for 10 min and supernatant protein concentration was measured using a BCA assay (Euroclone). Equal amounts of two different cell preparations of each cell line were pooled together.

3.1.2 Fractionation at protein level and in-gel digestion

A fraction of the total cell lysate of CFBE41o- was separated by One-Dimensional Gel Electrophoresis (1DGE). A total of 720 µg of the pooled lysate were separated (60 µg per 12 lanes) onto gradient 4-20% Criterion TGX Precast gel (Bio-rad laboratories Inc). The gel was fixed and colored with Novex Colloidal Blue Staining Kit (Thermo Sci.). Each lane was cut into 16 slices and in-gel digestion [152] with trypsin was performed.

Briefly, the gel was placed on a glass support to cut gel lanes into 16 slices. Using a scalpel, the gel slices were crumbled in small pieces of about 1-2 mm and placed in a 1.5-mL tube. As a first step, 100 µL of destaining solution (50% 50 mM ammonium bicarbonate / 50% acetonitrile) were added to each tube, incubated for 10 minutes under mild shaking, for the removal of the Coomassie dye. The destaining solution was removed and the destaining step was repeated as described before, in order to achieve a complete removal of dye. The reduction of disulfide bonds was obtained by incubating the samples at 56°C with 100 µL of 10 mM dithiothreitol in digestion buffer (50 mM ammonium bicarbonate in water, pH 8) for 1 hour. The reducing solution was discarded and the pieces of gel were washed with 100 µL of digestion buffer. The digestion buffer was removed and the gel pieces were incubated for 45 minutes in the darkness with 100 µL of alkylating solution (55 mM iodoacetamide in digestion buffer). After discarding the iodoacetamide solution, gel pieces were washed twice with 100 µL of digestion buffer for 10 minutes and then incubated for dehydration with Acetonitrile under shaking. The Acetonitrile was removed and the gel pieces were left evaporating for about 30 minutes under the hood. Each sample was incubated with 1 µg of trypsin dissolved in digestion buffer (enzyme concentration of trypsin solution 12.5 ng/µL) overnight at 37°C.

3.1.3 Protein enrichment with ProteoMiner kit

The Proteominer enrichment strategy (Biorad) was used [153], following the protocol suggested by the vendor. In short, protein concentration from CFBE41o- lysates was adjusted to 20 mg/mL by using Amicon Ultra centrifugal filter devices (3KDa cutoff, purchased from Millipore, USA). Proteominer medium was first activated with PBS (150 mM NaCl, 10 mM NaH_2PO_4, pH 7.4, 0.2 mL) and a total of 0.14 mL of protein solution was loaded and incubated for 2 hours. The column flow-through was removed and collected, along with two additional washes with 0.2 mL PBS. Proteins were then eluted eight times with 0.02 mL of 8 M urea, 5% CH3COOH, 2% CHAPS. In-solution digestion (Paragraph 3.1.4) was then performed on both wash and eluted fractions.

3.1.4 In-solution digestion

The volume corresponding to 50 µg of proteins sample was reduced with 10 µL dithiothreitol (DTT) 100 mM at 56 °C for 30 minutes, then alkylated with 30 µL of iodacetamide (IAA) 100 mM for 20 minutes in the dark. Protein content was then precipitated in cold (-20 °C) acetone overnight at -20 °C. To remove acetone, centrifugation at 15000xg for 30 minutes at 4 °C was performed. The supernatant was discarded and the pellet was dried under nitrogen stream before an overnight trypsin digestion at 37 °C in digestion buffer. The resulting peptides were dried under vacuum for further analysis.

3.1.5 Fractionation at peptide level

Tryptic peptides deriving from in-solution digestion (Paragraph 3.1.4) were fractionated offline with the high pH/low pH strategy [154]. This fractionation was performed using Oasis solid-phase extraction (SPE) columns, following the instructions provided by the vendor. Briefly, after the activation and equilibration of the column resin, the peptides were loaded on the column. After the desalting of the peptides, a total of 8 fractions of peptides were collected by eluting the sample with 500 µL of 0.1% TEA solution with increasing acetonitrile concentration (5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20% and 50%).

3.1.6 BE primary cell cultures and lysis

Human primary bronchial epithelial (BE) cells were obtained from individuals undergoing lung transplant. For this study, cells were obtained from four CF patients (BE86, BE91: patients with F508del/F508del genotype; BE29: patient with F508del/R553X genotype; BE148: patient with R1006C/M1V genotype) and four non-CF donors (BE61: patient with bronchiectasia; BE65: patient with idiopathic pulmonary fibrosis; BE147, BE121: patients with pulmonary emphysema). The collection and processing of human cells were approved by the local ethics committee, and written informed consent was obtained from all patients.

BE primary cells were cultured, transfected and then lysed in RIPA buffer added with complete protease inhibitor (Roche). Cell lysates were centrifuged at 15000xg at 4°C for 10 min; the supernatant was collected and the protein concentration was measured using a BCA assay (Euroclone). 50 µg proteins from each sample (both CF and non-CF) were pooled together for nanoLC-MS/MS analysis.

3.1.7 nanoLC-MS/MS analysis (DDA mode)

Dried tryptic peptides (from both CFBE41o- and BE primary cells, for a total of 40 samples) were dissolved in 150 µL of 3% acetonitrile (ACN) + 0.1% formic acid (FA) for mass spectrometry coupled to liquid chromatography (LC-MS/MS) analysis. The peptides were analyzed on a 5600+ TripleTOF instrument (SCIEX) coupled to a NanoAcquity LC system (Waters, Milford, MA, USA) and working in nanospray mode with a NanoSpray III ion source. The peptides were loaded and desalted on a trapping column (180 µm × 20 mm Acquity C18) for 4 min at 4.0 µL/min flow rate (1% ACN + 0.1% FA), and then separated on a PicoFrit C18 column (75 µm x 25 cm, from NewObjective Inc., Woburn, MA, USA). The peptides were eluted with a 2-h linear gradient at 300 nL/min (from 3% to 45% of ACN in water + 0.1% FA); the column was washed with 90% ACN and then equilibrated for 18 min to 3% ACN.

DDA (Data-Dependent Acquisition) spectra were collected over the two-hour gradient. Peptides with charge states 2+ to 5+ showing an intensity higher than 150 counts were selected as precursors for MS/MS acquisition. A survey spectrum (400-1250) was acquired for 250 ms, followed by 40 DDA MS/MS experiments (100-1500 *m/z*, 100ms accumulation time each).

3.1.8 Data analysis for protein identification

Raw DDA data from nanoLC-MS/MS analysis were analyzed with ProteinPilot software (SCIEX) using the Paragon algorithm [155]. All the DDA spectra were searched against a reviewed Homo sapiens database reporting 20303 proteins and downloaded as FASTA file from Uniprot (http://www.uniprot.org/proteomes/ UP000005640) on January 3rd, 2018. A False Discovery Rate (FDR) [156] analysis was performed against a decoy version of the same database and identifications at 1% FDR at spectrum level were retained and with a minimum threshold of 90% in peptide confidence.

3.1.9 Interaction maps

All the interaction maps present in this work were created using an open source tool available online called STRING (acronyms for Search Tool for the Retrieval of Interacting Genes/Proteins v.11.0, https://string-db.org/) [157].

The parameters used to detect possible interaction in the annotated, high-quality database of *Homo sapiens* were set in the "Data Settings" tab as follow:

- Meaning of network edges: evidence.
- Active interaction sources: Textmining, Experiments, Databases, Co-expression, Co-occurrence.
- Minimum required interaction score: high confidence (0.7).
- Display simplifications: disable structure previews inside network bubbles and hide disconnected nodes in the network.

With this tool, it is possible to obtain interaction maps showing the proteins as nodes connected by edges with different colors, corresponding to the different types of evidence used to show the interactions (**yellow** line indicates a textmining evidence; **pink** is used for experiments evidence; **light-blue** for databases; **black** for co-expression; **blue** for co-occurrence).

3.2 Proteomics profiling of BE primary cells (CF vs non-CF)

3.2.1 Sample preparation

BE primary cells deriving from lung transplant were cultured and lysed as described above (Paragraph 3.1.6). Protein content was digested following the in-solution digestion, already described (Paragraph 3.1.4).

3.2.2 nanoLC-MS/MS analysis (SWATH)

Dried peptides deriving from trypsin digestion were dissolved in 150 µL of 3% acetonitrile (ACN) + 0.1% formic acid (FA) for mass spectrometry coupled to liquid chromatography (LC-MS/MS) analysis. The peptides were analyzed on a 5600+ TripleTOF instrument (SCIEX) coupled to a NanoAcquity LC system (Waters, Milford, MA, USA) and working in nanospray mode with a NanoSpray III ion source. The peptides were loaded and desalted on a trapping column (180 µm × 20 mm Acquity C18) for 4 min at 4.0 µL/min flow rate (1% ACN + 0.1% FA), and then separated on a PicoFrit C18 column (75 µm x 25 cm, from NewObjective Inc., Woburn, MA, USA). The peptides were eluted with a 2-h linear gradient at 300 nL/min (from 3% to 45% of ACN in water + 0.1% FA); the column was washed with 90% ACN and then equilibrated for 18 min to 3% ACN.

For SWATH acquisitions, 1.66 µg of peptides were loaded on the chromatographic column. Precursor ion selection was done in the 400-1250 *m/z* range, with a variable window width strategy (7 to 50 Da). After a full range survey scan of 250 ms, 100 consecutive SWATH experiments (100-1500 *m/z*) were performed, each lasting 25 ms.

3.2.3 Data analysis for SWATH label-free quantification

The DIA spectra were searched against our modified Pan Human ion library [151], using only non-shared peptides. For the quantification, the following criteria were used: minimum peptide confidence 90%, 50 ppm maximum mass tolerance, 60 min maximum RT tolerance, 6 MRM transitions per peptide and modified peptides were not allowed.

Raw data were imported in MarkerView software and normalized with the most likely ratio (MLR) method [144]. The normalized data were pre-processed with logarithm weighting and Pareto scaling [158]. For statistical analysis, one-to-one comparisons were performed using an un-paired, 2 tails t-test, considering a p-value < 0.05 as significant. Moreover, according to the resulting t-value, the significantly dysregulated proteins were divided into either up- or down-regulated.

3.2.4 Pathway analysis

In this work, two different tools were mainly used for the pathway analysis: Reactome and GOrilla.

Reactome [159] is an open-source, manually curated and peer-reviewed pathway database. It provides bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge. The user can insert a list of proteins (generally the significantly up/down-regulated proteins, resulting from a proteomic experiment) and Reactome gives the user a graphical map and the information of the pathways in which the inserted proteins are involved in. The pathways are ranked according to the FDR value (probability corrected for multiple comparisons), calculated without considering the background of the sample.

Using the bioinformatics tool called ***GOrilla*** [160], the user can supply two lists of unranked proteins (the ranking of both lists does not affect results): the target and the background sets. The software is able to search for the pathways that are enriched in the target set compared to the background set using the standard Hyper Geometric statistics [161]. Selection of an appropriate background is therefore very important and could be challenging in high-throughput studies. In proteomics, the background list can be the complete set of proteins known to be expressed in an organ/tissue/cell line. This is very important since only a fraction of proteins are expressed in a given tissue [162] and only a subset of proteins can be quantified in a single proteomic experiment. In the present work, all the quantified proteins in each experiment were set as background list and the p. value threshold was set at 10^{-3}. The list of the significantly dysregulated proteins was used as target list in order to detect the pathways altered by those proteins.

3.3 Proteomics profiling of CFBE41o- cells after gene silencing

3.3.1 CFBE41o- cell culture, gene silencing and evaluation of CFTR rescue

For this study, CFBE41o- stably expressing F508del-CFTR and the halide-sensitive yellow fluorescent protein (HS-YFP) were used [148]. The gene silencing of selected proteins was performed by reverse transfection on CFBE41o- cells expressing F508del-CFTR and the HS-YFP using 10 nM (final concentration) siRNAs and lipofectamine 2000 as transfection agent.

In this experiment, five different F508del-CFTR expressing CFBE41o- cells groups, in five independent biological replicates, were prepared and incubated for 48 hours: the control group (treated with dimethyl sulfoxide, DMSO) and groups treated with siRNA against four different proteins:

- ✓ ***RNF5*** (a ubiquitin ligase involved in recognition of CFTR folding defect [148]).
- ✓ ***FAU*** (a ribosomal protein whose biological functions are still largely unknown [163]).
- ✓ ***LRRC59*** (a nuclear protein known to be involved in nuclear import of exogenous fibroblast growth factor 1 [164]).
- ✓ ***PHF15*** (an E3 ubiquitin-protein ligase involved in protein ubiquitination pathway [165]).

Evaluation of CFTR protein expressed at the plasma membrane was done by performing a cell surface biotinylation assay, followed by immunoprecipitation of biotinylated proteins and western blot analysis [166] on total cell lysates. With this assay, it is possible to look for changes in band C (fully glycosylated mature CFTR) and compare it to band B (core-glycosylated immature CFTR). A comparison between lysates of CFBE41o- cells expressing F508del-CFTR following different treatments (siRNAs and/or DMSO) and lysates of CFBE41o- cells expressing wild type CFTR was performed.

3.3.2 YFP-based functional assay

YFP-based assay was performed to evaluate the effect of F508del-CFTR functional rescue after gene silencing. The assay was performed following a well-established protocol [166], using CFBE41o- cells that express F508del- CFTR together with a yellow-fluorescence protein (YFP)-based halide sensor.

At the time of the assay, cells were washed with PBS and then incubated for 25 min with 60 µL of PBS plus forskolin (20 µM). For CFTR activity determination, cells were transferred to a microplate reader (FluoStar Galaxy; BMG Labtech, Offenburg, Germany). The plate reader was equipped with high-quality excitation (HQ500/20X: 500 ± 10 nm) and emission (HQ535/30M: 535 ± 15 nm) filters for YFP (Chroma Technology). Each assay consisted of a continuous 14-s fluorescence reading with 2 s before and 12 s after injection of 165 µL of an iodide-containing solution. Data were normalized to the initial background-subtracted fluorescence. To quantify the rate of fluorescence change, the final 11 s of the data for each well were fitted with an exponential regression function to extrapolate initial slope (dF/dt). Statistical significance was assessed by parametric ANOVA, followed by Dunnett's multiple post-hoc test, which is generally used to compare each of a number of treatments with a single control.

3.3.3 Sample preparation and data analysis

CFBE41o- cells after the gene silencing and the evaluation of CFTR rescue, were lysed and their protein content was digested as already described (Paragraph 3.1.4). The corresponding tryptic peptides were analyzed following the nanoLC-MS/MS workflow for SWATH quantitative proteomics (Paragraph 3.2.2, Paragraph 3.2.3).

The corresponding raw data were imported in MarkerView software and normalized with the most likely ratio (MLR) method [144]. The normalized data were subsequently scaled with logarithm weighting and Pareto scaling [158]. For statistical analysis, one-to-one comparisons were performed using an un-paired, 2 tails t-test, considering a p-value < 0.05 as significant. Moreover, according to the resulting t-value, the significantly dysregulated proteins were divided into either up- or down-regulated.

3.3.4 Selection of target candidates

To select new target candidates an extensive literature search was conducted to study the role of the proteins (and pathways) altered by the gene silencing, using the free search engine PubMed [167], [168].

Four parameters were then taken into account to select appropriate modulators:

1) **Biological activity**: the modulator must efficiently act on its target(s).
2) **LogP**: this parameter is the index of the lipophilicity of a molecule and it was taken into account in order to select modulators that were soluble in the medium used to culture CFBE41o- cells.
3) **Safety**: the modulators should have no reported adverse effects.
4) **Availability**: the modulators to be tested must be commercially available.

PubMed was used to study the role of the proteins that were significantly altered by the gene silencing and how they could be related to CFTR rescue. Moreover, two databases were interrogated to select the appropriate modulators that showed all the criteria described above: ChEMBL [169] and DRUGBANK [170].

PubMed, Uniprot [170] and REACTOME [159] database were used to study proteins involved in the pathways significantly altered by the gene silencing. Among these proteins, CFTR interactors were searched using STRING (refer to Paragraph 3.1.9) in order to select target candidates that were searched into ChEMBL database to find modulators fulfilling all the criteria requested for further experiments.

3.4 Lipidomics profiling of CFBE41o- after incubation with drugs

3.4.1 Cell cultures, incubations with drugs and CFTR rescue evaluation

Selected CFBE41o- cells over-expressing F508del-CFTR were cultured until confluence and treated with vehicle alone (DMSO) or with the following drug (or combination of drugs) concentrations: VX-809 (3 µM), VX-661 (10 µM), VX-770 (5 µM), VX-445 (3 µM), VX-809/VX-770, VX-661/VX-770, VX-445/VX-770, VX-661/VX-445, VX-661/VX-445/VX-770. To ensure reproducibility, six independent experiments were performed using 5 million cells per condition. Cells were harvested in PBS, pelleted, lysed (Paragraph 3.1.6) and CFTR rescue was measured (Paragraph 3.3.2).

3.4.2 Lipid extraction

Cells pellets were resuspended with H_2O (50 µL), transferred to glass vials, and added with 1 mL isopropanol spiked with C17 Ceramide (1 µM) as internal standard. The samples were then vortexed for 10 minutes and sonicated for 10 minutes at room temperature (RT). After a centrifugation step (at 20000 x g for 20 minutes at RT), the first supernatant was carefully collected (Named A) and the pellets were re-extracted with 100 µL of methyl-tert-butyl ether (MTBE) /methanol (50:50 v/v). After vortexing for 10 minutes and sonicating for 10 minutes at RT, another centrifugation step (at 20000 x g for 20 minutes at RT) was performed and the second supernatant was collected (Named B). Supernatants A and B were then pooled and dried under nitrogen stream. All the samples, including procedure blanks, were extracted following this protocol.

3.4.3 LC-MS/MS analysis for untargeted lipidomics

The dried lipids extracts were dissolved in 200 µL of methanol/chloroform (9:1, v/v) for LC-MS/MS analysis. The whole sample set (a total of 64 samples) was split into 4 batches and analyzed by high-resolution LC-MS/MS, using a well-established method for untargeted lipidomics [171]. For each batch quality control (QC) samples, prepared pooling together 5 µL for each batch, were prepared and analyzed together with the samples within each batch.

The lipidomics analysis was performed on an ACQUITY UPLC system coupled to a Synapt G2 QTOF high-resolution mass spectrometer (Waters, Milford, MA, USA), acquiring both in positive (ESI+) and negative (ESI−) ion modes. Lipids separation was performed with a reverse-phase CSH C18 column (1.7 µm internal diameter; 2.1x50 mm, Waters). Eluents were acetonitrile/water (60:40 v/v) with ammonium formate 10 µM (Eluent A) and isopropanol/acetonitrile (90:10 v/v) with ammonium formate 10 µM (Eluent B). Injection volume was 3 and 5 µL for ESI+ and ESI- respectively. The flow rate was set on 0.450 mL/min, the column was kept at 50 °C. Samples were eluted with the following gradient profile: 0.0 -1.0 min 10 % B; 1.0 – 4.0 min 10 to 60 % B; 4.0 – 8.0 min 60 to 75 % B; 8.0 – 8.5 min 75 to 100 % B; 8.5 – 10.0 min 100 % B and 10.0 – 10.1 back to 10 % B. The column was then reconditioned for 1.9 min. The total run time was 12 min. Scan range was set from 50 to 1200 *m/z*. Cone voltage was set at 35 V. Source temperature was set to 90 °C, desolvation gas and cone gas flows were set to 800 and 50 L/h respectively, desolvation temperature was set to 400 °C. Data were acquired in MSE mode, alternating MS and MS/MS scans [172]. The scan time was set to 0.3 s, low collision energy was set to 4 eV and 13 high collision energy was ramped from 25 to 45 eV. Leucine enkephalin (2 ng/mL) was infused as lock mass for real-time spectra recalibration. Masslynx software (from Waters) was used for data acquisition.

3.4.4 Targeted ceramide and LysoPC analysis

To confirm the findings of untargeted lipidomics, ceramides were also quantified in a targeted experiment, by using a triple-quadrupole instrument (Xevo-TQMS, Waters) coupled to an ACQUITY UPLC (Waters) chromatographic system, following a dedicated method [173]. Three Lyso PCs were quantified by using the same general microLC-MS/MS method and adding the three corresponding MRM experiments (496 → 184, 524 → 184 and 52 →184 for Lyso PCs 16:0, 18:0 and 18:1 respectively). The collision energy was set to 25 eV and cone voltage was set to 30V.

3.4.5 Evaluation of cell susceptibility to pro-apoptotic stimuli

CFBE41o- cells stably expressing F508del-CFTR were plated at low density (10,000 cells/well) on proper 96-well plates. After 24 hours, the pharmacological treatment was performed in the absence or in the presence of different concentrations (1 – 10 µM range) of Etoposide to induce cell apoptosis. The following day plates were washed three times with DPBS to remove dead cells, cell nuclei were counterstained with Hoechst 33342 and propidium iodide to visualized total and apoptotic cells, respectively. Plates were imaged with a 10X air objective using the Opera Phenix (PerkinElmer) high-content screening system. The excitation of Hoechst 33342 signal was at 405 nm and the emission at 435-480 nm. The propidium iodide signal was excited at 560 nm and the emission was measured at 570-630 nm. Data are expressed as means ± SEM, n = 8. Reproducibility of results was confirmed by performing three independent experiments. Statistical significance was tested by parametric ANOVA, followed by the Dunnett'st multiple comparisons test.

3.4.6 Data analysis and statistics

The observed features for both ESI+ and ESI- polarities were extracted from the RAW data, integrated and re-aligned over all the runs by using the Markerlynx software (Waters, Inc.). The following parameters were set: mass range from 200 to 1200 *m/z*, retention time (RT) range from 1 to 10 minutes, minimum intensity 4000 ion counts, *m/z* tolerance for extracted ion current 0.03 *m/z*, RT tolerance 0.03 min. Background ions observed in procedure blank and blank runs above the same threshold were automatically excluded by the software, using a background exclusion list. Exogenous d18:1/17:0 ceramide was used as internal standard, and it was observed as $[M+H_2O+H]^+$ and $[M+HCOO]^-$ adducts for ESI+ and ESI- ion modes respectively with RT 5.47±0.03 min. The *m/z* values of the drugs were excluded from the peak picking process, even though the cell pellets were previously washed before lipid extraction.

The samples from all the four acquisition batches were realigned using this ceramide as reference. The obtained datasets for ESI+ and ESI- were then merged in a final (1571 features X 42 samples) Excel file and analyzed using Metaboanalyst [141] web-based software. The missing values were automatically replaced with the minimum value of each given feature in the corresponding column. The dataset was then normalized by the sum of all the observed features, log transformed and Pareto scaled [158] for multivariate data analysis. Both the unsupervised (Principal Component Analysis – PCA) and supervised (Partial Least Squares Discriminant Analysis – PLS-DA) analysis methods were used, with the aim to identify changes in the global lipidomic profiles. The cross validation of PLS-DA analysis was performed by using the leave-one-out algorithm [174], searching over 5 principal components. Variable Importance in Projections (VIP) score for PLS-DA was calculated over component 1. Cluster analysis was done using heatmaps, by applying Euclidean distance measurement and Ward algorithm [141] to normalized data.

Correlation analysis among the features was performed using Pearson r as distance measure [175]. Pathway analysis was performed on annotated data, using their HMDB [176] identifier. Hypergeometric test was used for overrepresentation analysis and Relative-betweenness Centrality was used for Pathway Topology Analysis on the KEGG [177] pathway library for Homo Sapiens.

For binary comparisons, Volcano plot analysis was used to select features based on their observed fold change between the treated Vs control groups (≥1.5) and their FDR adjusted p-value based on a t-test ($p<0.05$). For univariate data analysis, both 2 tailed unpaired t-tests and one-way analysis of variance (ANOVA) were performed using GraphPad Prism (GLS Biotech). Dunnett'st post –hoc test was used to compare all groups with each other.

3.4.7 Features annotation

The relevant features, selected from VIP scores and t-tests from Volcano plot analysis, were manually annotated searching them against the LipidMaps database [178], considering only the following possible adduct species: $[M+H]^+$, $[M-H_2O+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ for ESI+ and $[M-H]^-$, $[M+HCOO]^-$ for ESI-, allowing 0.01 *m/z* as maximum allowed tolerance for both polarities. The resulting list of proposed matches was then manually scrutinized, taking into account accurate mass and MS/MS data, beside the consistency of observed adduct species and retention time [179]. None of the proposed annotations was confirmed with the use of an authentic analytical reference standard, so, based on current guidelines for reporting data in metabolomics [178], lipid IDs should thus be considered putative.

3.5 LOPIT-DC experiments

These experiments were conducted in Cambridge, where I spent five months, hosted by Prof. Lilley's lab.

LOPIT-DC [87] workflow was applied to the CFBE41o- cells to monitor the protein localization changes associated with CFTR rescue in order to pinpoint those proteins which could be detrimental (or beneficial) to the functional rescue of the channel. The entire experimental procedure (already described in Paragraph 1.2.1.1) is summarized in Figure 25.

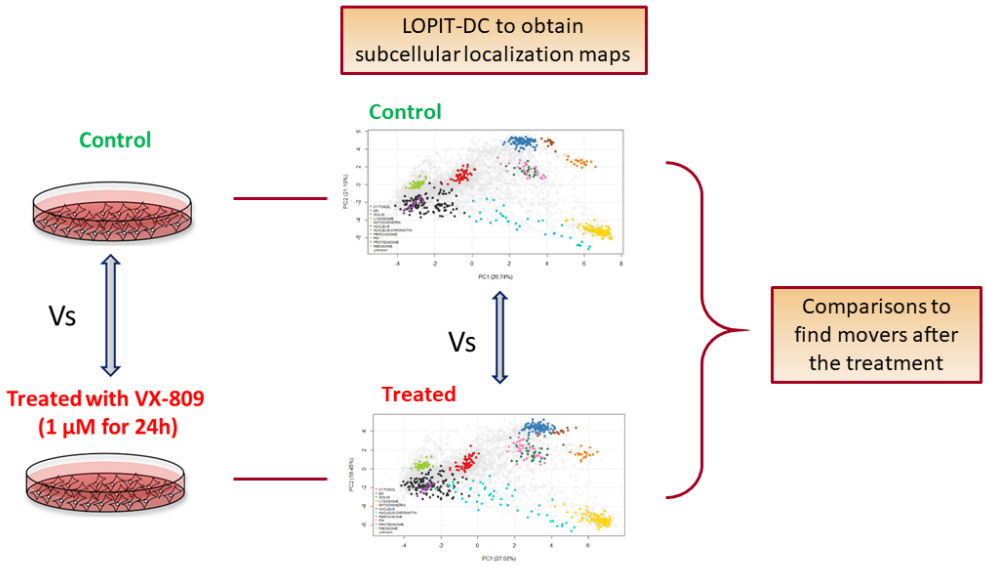


Figure 25: LOPIT-DC workflow applied to the CFBE41o- to detect proteins with different subcellular localization after the treatment with VX-809.

3.5.1 Cell cultures, VX-809 treatment and ultracentrifugation steps

Selected CFBE41o- cells over-expressing F508del-CFTR were cultured until confluence as already described (Paragraph 3.1.1). For this experiment, 70 million cells per sample are needed. Three independent biological replicates were prepared, processing together the treated with its control in order to decrease technical variability. For each replicate, the treatment was performed incubating the cells with medium containing either the VX-809 1 µM or the vehicle (DMSO, used as control) for 24 hours.

Cells were harvested with TrypLE Express Enzyme (Gibco) and pelleted at 300xg for 7 minutes at 4°C. Cell pellets were suspended in 5 mL of a gentle lysis buffer (0.25M sucrose, 10mM HEPES pH 7.4, 2mM EDTA, 2mM magnesium acetate, protease inhibitors). Cells were then lysed using a pair of 1 mL syringes in a ball-bearing homogenizer (Isobiotec): each aliquot of cells suspension was passed through the homogenizer chamber 30 times using a 18 µm ball-bearing clearance, keeping the homogenizer on ice.

Cell lysates were processed following the LOPIT-DC workflow [87], where after every centrifugation step, the supernatant is collected and subjected to the following centrifugation step. The speeds and times for each step are listed in Table 1.

Fraction name	Speed (xg)	Time (min)
Whole cell removal	200	5
P1	1000	10
P2	3000	10
P3	5000	10
P4	9000	15
P5	12000	15
P6	15000	15
P7	30000	20
P8	79000	43
P9	120000	45
SN	120000 SN	

Table 1: Speed and times of the differential centrifugation steps of the LOPIT-DC workflow. P=pellet and SN=supernatant.

After the removal of unlysed cells (essential to prevent alterations in the downstream analysis), a total of ten fractions were collected for each sample. The protein content of the supernatant fraction was precipitated with incubation with chilled acetone (8xV) at -20 °C overnight.

The ten pellets were solubilized using different volumes of membrane solubilizing buffer (MSB, 8 M Urea, 0.2% SDS, 50 mM HEPES, pH 7.4). Protein content was measured using the Pierce BCA protein assay kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

3.5.2 Evaluation of organelles separation through Western-Blot (WB)

2 µg of proteins from each fraction was separated onto gradient 4-20% Mini-PROTEAN TGX Precast gel (Bio-rad laboratories Inc.) and then transferred onto a Polyvinylidene fluoride membrane (Bio-rad laboratories Inc.). The membranes were blocked and cut into strips to be incubated with the primary antibodies. Table 2 summarizes the antibodies used for each organelle.

Organelle	Antigen	Gene name	MW (kDa)	Source	Supplier	Dilution for WB
Cytosol	Alfa-Enolase	ENO1	47	Rabbit	Cell signaling	1:1000
Golgi	Syntaxin-6	STX6	32	Rabbit	AbCam	1:1000
Plasma membrane	Alfa1 Na+/K+ ATPase	ATP1A1	100	Rabbit	AbCam	1:5000
ER	Calreticulin	CARL	55	Rabbit	AbCam	1:1000
Chromatin	Histone H2A	HIST1H2AG		Rabbit	AbCam	1:1000
Endosome	Early endosome antigen 1	EEA1	170	Rabbit	AbCam	1:1000
Lysosome	Lysosome-associated membrane glycoprotein 1	LAMP1	100	Rabbit	Cell signaling	1:1000
Nuclear–non chromatin	Fibrillarin	FBL	37	Rabbit	Cell signaling	1:1000
Nuclear envelope	Prelamin-A/C	LMNA	70	Rabbit	AbCam	1:1000
Mitochondria	Cytochrome c oxidase subunit 4 isoform 1 (COX 1V)	COX4I1	17	Rabbit	Cell signaling	1:1000

Table 2: List of antibodies for evaluating organelle-specific marker proteins.

After the incubation with the secondary antibody and addition of the Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) solution, the membranes were exposed to Fuji films in the dark to be used for the evaluation of the separation of the organelles.

3.5.3 Protein digestion and TMT labelling

The volume corresponding to 50 µg of proteins from each fraction was transferred into new Eppendorf Lo-bind tubes and in-solution digestion was performed following a protocol already described (Paragraph 3.1.4).

The tryptic peptides were labelled with TMT isobaric tagging reagents (Thermo Fisher Scientific) according to manufacturer's protocol. Since each TMT kit can be used to tag up to 100 µg of proteins, half of the TMT is used for each sample (50 µg of proteins). For this reason, each kit was split in half to be used for each replicate (half for the control and half for the treated). The TMT tags were equilibrated at room temperature and each tag was dissolved in 82 µL of acetonitrile. After vortexing and centrifuging 2 minutes at 2000xg, 41 µL of TMT tag solution was added to label each fraction, following the scheme showed in Figure 26, for two hours on a shaker at room temperature.

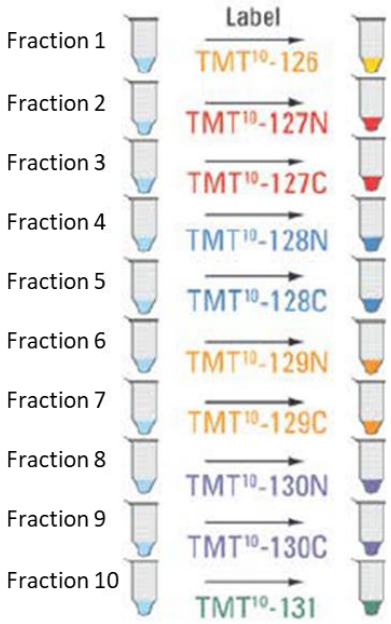


Figure 26: Scheme of the TMT labelling tags used for each fraction of the LOPIT-DC workflow.

The labelling reaction was quenched by adding 8 µL of 5% (w/v) hydroxylamine and incubated for 45 minutes at room temperature on a shaker. Another quenching step was performed by incubation with 100 µL of MilliQ water for 30 minutes at room temperature on a shaker. After the TMT labelling, all the fractions were pooled together and dried under vacuum.

3.5.4 Peptide fractionation and LC-MS/MS analysis

Dried TMT-labelled peptides were resuspended in 1.8 mL of 0.1% TFA (trifluoroacetic acid) and 600 µL of this solution was used for the peptide fractionation using the Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific) according to manufacturer's protocol. In summary, the columns were activated and equilibrated before loading the samples. The peptides were then washed and eluted with different solvents, described in Table 3.

Fraction	% ACN in water
Wash	5.0
1	10.0
2	12.5
3	15.0
4	17.5
5	20.0
6	22.5
7	25.0
8	50.0

Table 3: Composition of the eluents used for peptide fractionation.

In order to decrease the number of the samples for downstream MS analysis (8 fractions of peptides per sample, 6 samples in total), the resulting fractions were combined into 4 fractions per sample (Fraction 1+5; Fraction 2+6; Fraction 3+7; Fraction 4+8).

All MS runs were performed on an Obritrap Eclipse Tribrid Mass Spectrometer coupled to a Dionex Ultimate TM 3000 RSLnano system (Thermo Fisher Scientific). Each of the fractionated samples was dissolved in 60 µL of 0.1% formic acid and 2 µL were injected onto a micro precolumn (300-µm x 5 mm, C18 PepMap 100, 5-µm particle size, 100-Å pore size, Thermo Fisher Scientific) for the trapping phase (3 minutes). After the switching of the valve from load to inject, the peptides were separated on a 200 cm µPAC™ column (PharmaFluidics). The nanoLc gradient was set as follows: from 3 to 40% of ACN in water + 0.1% FA at 300 nL/min for 340 minutes. The column was then washed using 95% of ACN + 0.1% FA for 9 minutes and re-equilibrated for 40 minutes. Total run time was 390 minutes.

The MS workflow parameters were set as follows using the Method Editor in XCalibur (Thermo Fisher Scientific) for the SPS-MS3 acquisition method: Detector type: Orbitrap—Resolution: 120,000—Mass range: Normal—Use quadrupole isolation: Yes—Scan range: 400–1,500—RF lens: 30%—AGC target: 40000—Max inject time: Auto—Microscans: 1—Data type: Profile—Polarity: Positive—Monoisotopic peak determination: Peptide—Relax restrictions when too few precursors are found: Yes—Include charge state(s): 2–6—Exclude after n times: 1—Exclusion duration (s): 70—Mass tolerance (p.p.m.): Low: 10; high: 10— Exclude isotopes: Yes—Perform dependent scan on single charge state per precursor only: Yes—Intensity threshold: 5.0e3—Data-dependent mode: Cycle time —Number of scan event types: 1—Scan event type 1: No condition—MSn level: 2— Isolation mode: Quadrupole—Isolation window (m/z): 0.7—Activation type: CID —CID collision energy (%): 35—Activation Q: 0.25—Detector type: Ion trap—Scan range mode: Auto—m/z: Normal—Ion trap scan rate: Rapid—AGC target: 1.0e4— Max inject time (ms): 35—Microscans: 1—Data type: Centroid—Mass range: 400–1500—Exclusion mass width: m/z: Low: 0; high: 0—Reagent: TMT— Precursor priority: Most intense—Scan event type 1: No condition—MSn level: 3— Isolation mode: Quadrupole—Isolation window (m/z): 0.7—Activation type: HCD —Collision energy (%): 65—Detector type: Orbitrap—Scan range mode: Define—m/z: 100-500—AGC target: 1.0e4—Max inject time (ms): 105—Microscans: 1—Data type: Centroid.

An electrospray voltage of 2.1 kV was applied. The mass spectrometer was operated in positive ion data-dependent mode for SPS-MS3 with the real-time (RT) search using the Homo sapiens Swiss proteome (downloaded 09/04/2018 as FASTA file). The parameters for the RT-search were: ezyme: trypsin; max variable mods: 2; max missed cleavage: 1; Static modification: carbamidomethylation on cysteine and TMT-10plex on lysine and peptide *N*-terminus; variable modifications: oxidation of methionine.

3.5.5 Data analysis and statistics

Raw files were processed with Proteome Discoverer version v2.3 (Thermo Fisher Scientific). The acquired MS/MS spectra were searched against the Homo sapiens database (canonical and isoform, 42,118 sequences, downloaded on 04/11/2016) using Proteome Discoverer with Mascot and SequestHF algorithms. The research parameters were set as follows: precursor mass tolerance: 10 ppm; fragment mass tolerance: 0.6 Da; enzyme used for digestion: Trypsin; maximum missed cleavage: 2; fixed modifications: carbamidomethylation of cysteine and TMT10plex tagging of lysine and peptide N terminus for TMT labelled samples; dynamic modifications: oxidation of methionine and deamidated asparagine and glutamine. Percolator node was used for false discovery rate estimation and only rank 1 peptide

identifications of high confidence (FDR < 1%) were accepted. TMT reporter values were assessed through Proteome Discoverer v2.3 using the Most Confident Centroid method for peak integration and integration tolerance of 20 p.p.m. Reporter ion intensities were adjusted to correct for the isotopic impurities of the different TMT reagents (according to the manufacturer specifications for the respective batch number).

Percolator was used to assess the false discovery rate (FDR) and only high-confidence peptides were retained (<1%FDR). Additional data reduction filters were: peptide rank= 1 and ion score >20. Quantification at the MS3 level was performed within the Proteome Discoverer workflow using the centroid sum method and an integration tolerance of 2 mmu. Isotope impurity correction factors were applied. Protein grouping was carried out according to the minimum parsimony principle and the median of all sum-normalised PSM ratios belonging to each protein group was calculated as the protein group quantitation value. Only proteins with a full reporter ion series were retained.

The quantified proteins were analyzed using the R Bioconductor [180] packages MSnbase and pRoloc following a well-established protocol for LOPIT datasets [181].

Briefly, each sample was normalized, the missing values removed and then combined to obtain a map of the organelles for each condition.

A list of 1590 manually-curated marker proteins (defined as proteins known to localize only in one specific organelle) was used to pinpoint 11 subcellular localizations: cytosol, ER, proteasome, Golgi, lysosomes, mitochondria, nucleus, chromatin, ribosome, peroxisomes and plasma membrane. To predict protein localization of the proteins quantified both in the controls and the treated samples, a support vector machine (SVM) classifier was used, following a protocol available in literature [181]. To detect proteins that changed their localization after the treatment with VX-809, TAGM (t-augmented Gaussian mixture) statistical analysis was performed, following a well-established workflow available in literature [89].

4 Results

4.1 Creation of the ion library for CF research

As mentioned in the introduction, for SWATH proteomics it is very important to setup a reliable ion library specifically dedicated and optimized for the proteome under investigation. This ensures the best accuracy and specificity in protein quantification. Both immortalized (CFBE41o-) and BE primary cells were then analyzed to create an ion library highly specific for CF research. During sample preparation, two kinds of fractionation methods were applied in order to deepen the coverage of the proteome.

The ion library (IL) deriving from the analysis of CFBE41o- reports assays for 5627 proteins and it only reports unique and unmodified peptides (peptides not shared, and not carrying PTMs other than cysteine carbamidomethylation and oxidation on methionines). For human samples, the PanHuman IL is the most complete ion library available so far and it derives from the analysis of different types of cells and tissues [150]. Despite its outstanding quality, the PanHuman IL lacks assays for several proteins essential to study CF (such as proteins involved in CFTR trafficking and CFTR itself). For this reason, we compared our new IL with the PH one in order to detect those proteins highly specific for CF that are not present in the PanHuman IL.

After comparing the CFBE41o- ion library and the PH, we identified 124 proteins that are specific for the human bronchial epithelium. Among these proteins, CFTR is present, together with some of its known interactors, such as RNF185, demonstrating the high specificity of this new ion library to study CF.

The IL of the BE primary cells, created exactly as the CFBE41o- cells one, reports assays for 3805 proteins.

A comparison of the CF-specific ion libraries with the PH one was then performed, in order to understand the number of proteins to be added. The results of this comparison are shown in Figure 27.

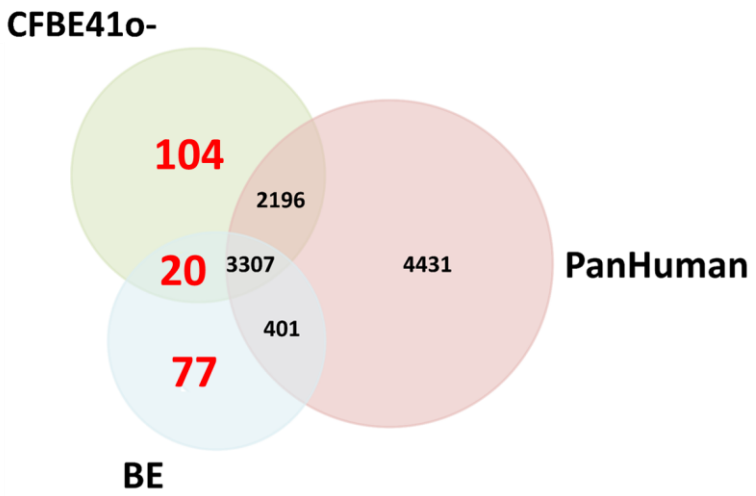


Figure 27: Overlap analysis comparing the proteins present in the PanHuman, the BE and CFBE41o- cells libraries. Figure taken from Braccia C., SWATH label-free proteomics for cystic fibrosis research [151].

The assays corresponding to CF-specific proteins (124 and 77 for CFBE41o- and BE primary cells, respectively) were manually added to the PanHuman ion library, thus producing a new and improved ion library, now reporting assays for 14683 proteins. These proteins include CFTR and other proteins very relevant for CF research, like RNF185, RNF5, DLG4 and STX1A. These molecules are known CFTR interactors, as reported in the STRING database (see Figure 28).

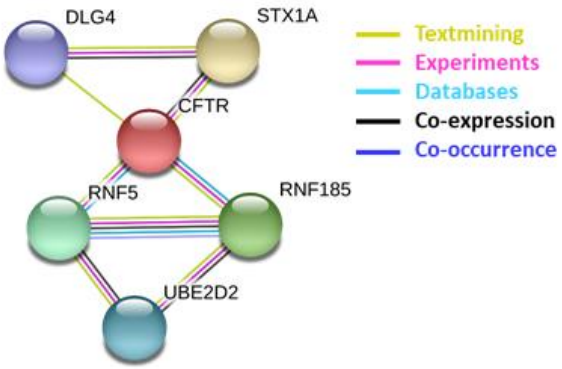


Figure 28: CFTR interactors among the 124 proteins identified only in the CFBE41o-. Figure taken from Braccia C., SWATH label-free proteomics for cystic fibrosis research [151].

4.2 Proteomics profiling of BE primary cells (CF vs non-CF)

To demonstrate the efficacy of the new ion library, a quantitative proteomics analysis was performed to monitor the changes in protein expression induced by CF pathology on BE cells directly obtained from CF subjects and non-CF controls, beyond the intrinsic inter-individual variability. To perform this analysis, the BE cells described above (Paragraph 3.1.6) were used: the proteome of four primary cultures of BE cells obtained from four CF patients was compared to the proteome of four non-CF subjects, used as control.

As a test, SWATH data were then acquired and searched against different ion libraries. As shown in Table 4, an increase in the number of quantified proteins was achieved with the use of the modified PH ion library.

Ion library	Total quantified proteins	Upregulated in CF	Downregulated in CF
PanHuman	4042	92	59
PanHuman & CFBE	4054	94	60
PanHuman, CFBE & BE	4074	94	60

Table 4: Number of quantified proteins from total BE lysates using different ion libraries. For each dataset, the number of proteins significantly altered by CF is shown (results from unpaired, two-tails t-test CF vs non-CF, 4 samples each).

By using the proteins quantified with the new version of the PH IL, a t-test (un-paired, two-tails t-test CF vs non-CF, 4 samples each) was then performed and 154 proteins significantly (p. value < 0.05) dysregulated by CF pathology (94 upregulated and 60 downregulated) were identified. As expected, given the low natural levels of expression, CFTR was not quantified in the primary BE dataset.

The interaction maps created with STRING (Paragraph 3.1.9) show that, among the proteins significantly up-regulated by CF, two proteins (PMSA4 and TJP1) are known CFTR interactors, as shown in Figure 29.

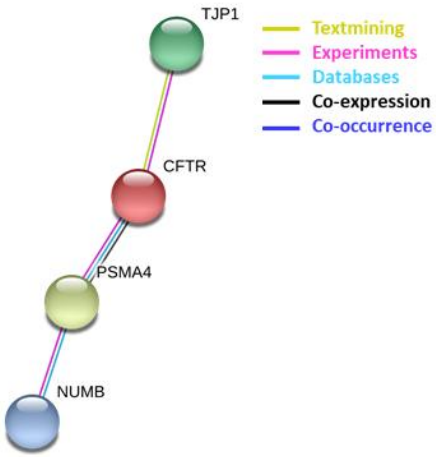


Figure 29: Detail of the CFTR cluster from the STRING analysis of the significantly up-regulated proteins by CF in BE cells. Since CFTR was not quantified, it was manually added to the dataset in STRING.

Proteasome subunit alpha 4 (PMSA4) plays a key role in the ATP-dependent degradation of ubiquitinated proteins [182], and in the REACTOME pathways database [159] it is directly associated to the degradation of ubiquitinated CFTR (pathway: R-HSA-5678895). Tight junction protein 1 (TJP1) is a known direct interactor of CFTR [106] and based on REACTOME, both proteins are involved in membrane trafficking (pathway: R-HSA-199991).

Among the up-regulated proteins detected by this experiment, there is a cluster of mitochondrial ribosomal proteins (MRPS22, MRPS15, MRPL44, MRPL40 and RPL37A) and proteins belonging to coatomer complex [183], known to regulate the Golgi-endoplasmic reticulum (ER) protein trafficking [184] (COPB1, COPE, ARF5 and GBF1), as shown in Figure 30.

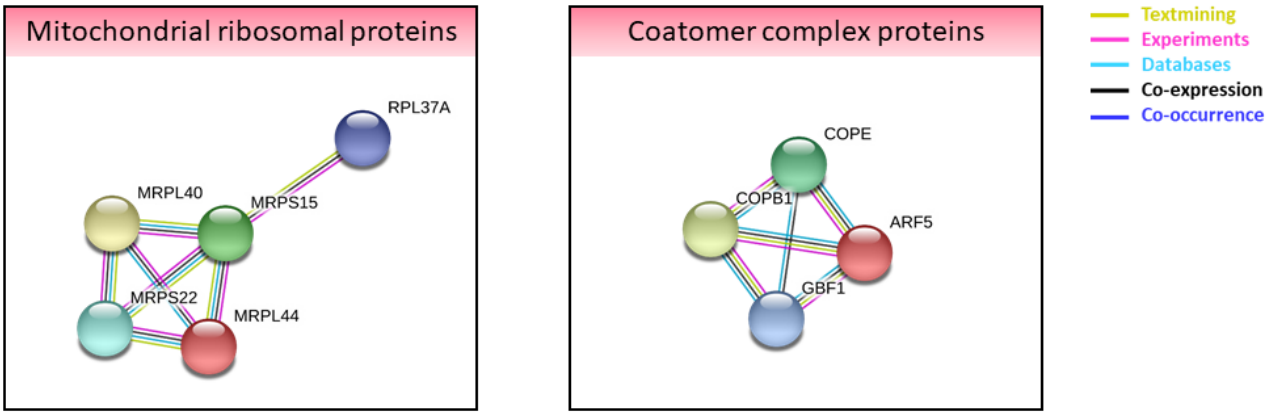


Figure 30: Two clusters of proteins (detected as up-regulated by CF) involved in protein trafficking.

Among the proteins significantly down-regulated by CF, one CFTR interactor was quantified (Figure 31).

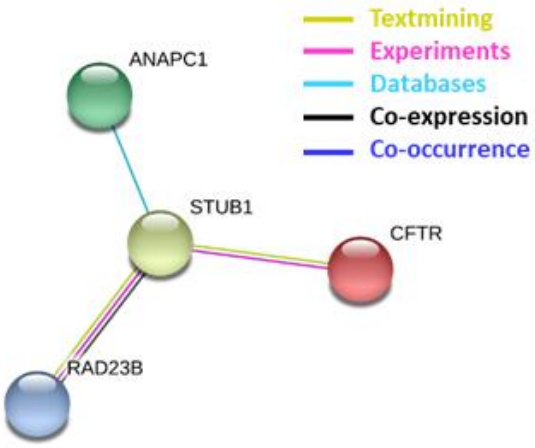


Figure 31: Detail of the CFTR cluster from the STRING analysis of the significantly down-regulated proteins by CF in BE dataset. Since CFTR was not quantified, it was manually added to the dataset in STRING.

E3 ubiquitin-protein ligase CHIP (STUB1), which has a crucial role in CFTR maturation [185], is downregulated in the dataset. As shown in Figure 31, STUB1 protein interacts with other two proteins that are downregulated in the BE of CF patients: anaphase-promoting complex subunit 1 (ANAPC1) and the multiubiquitin receptor RAD23 [186]. Together with STUB1 protein, ANAPC1 and RAD23 proteins are involved in the ubiquitin-mediated proteolysis pathway reported in the KEGG database [187].

The downregulation of STUB1 protein is apparently in contrast with previous literature (although obtained on totally different in vitro cell models, such as HEK293), where the decrease of CFTR is normally associated with CHIP upregulation [185].This event might, thus, be specific for BE and warrants further studies or larger patients cohorts.

Two proteins correlated to CFTR activity (PP1CA and PPP1R12C, two type-1 protein phosphatases) were detected among the downregulated proteins. While type-2 protein phosphatases are known to be associated with CFTR [188], type-1 phosphatase has more recently been demonstrated to regulate anion exchange in synergy with CFTR [189]. Very recently, S-nitrosation of type-1 phosphatase has been proved to regulate ciliary activity [190] and in-situ production of nitrosothiols (by S-nitrosoglutathione reductase inhibition) is sought as a clinical strategy to improve CFTR maturation [191].

4.3 Proteomics profiling of CFBE41o- after gene silencing

After demonstrating the efficiency of our new IL in detecting changes of proteins highly specific for CF in bronchial epithelium primary cells, we used our new tool also to quantify the proteins of CFBE41o- cell lysates in order to detect those proteins changing their expression after the gene silencing of four proteins (FAU, LRRC, PHF, RNF5, Paragraph 3.3.1).

The LC-MS/MS analysis of the CFBE41o- cells after the gene silencing of four proteins (Paragraph 3.3) allowed the quantification of 4010 proteins in all the samples. The unsupervised multivariate analysis was performed and the resulting PCA score plot is showed in Figure 32.

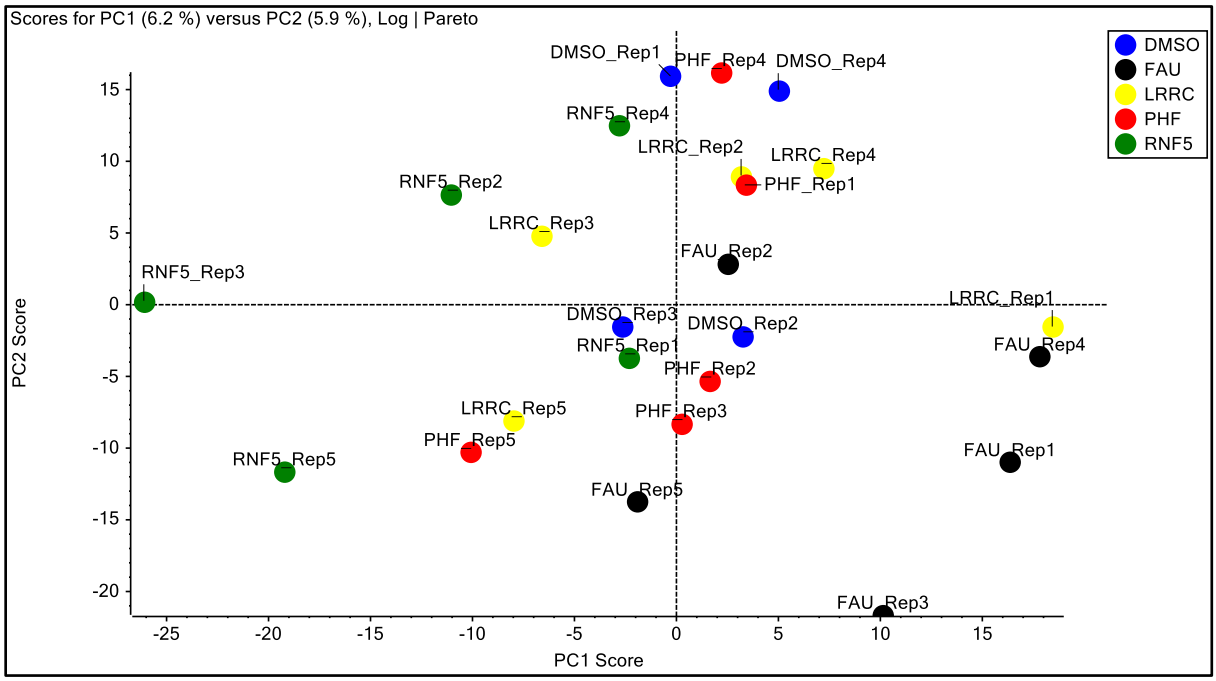


Figure 32: PCA score plot of all the samples after the gene silencing of the four proteins described above (Paragraph 3.3.1). For the multivariate analysis, the normalized data were pre-processed with logarithm weighting and Pareto scaling (Paragraph 3.3.3).

From the PCA score plot, it is possible to notice that none of the five groups is clearly separated from all the other groups (the distance between samples within the same group is higher than the distance of different groups) and that the total variance expressed by both PC1 and PC2 is 12.1%. (PC1=6.2% and PC2=5.9%). This suggests that the gene silencing performed does not have a huge impact on total protein expression.

To select the putative new targets, binary comparisons (Paragraph 3.3.4) were then performed to detect the significantly altered proteins by each gene silencing. Two strategies were followed to find new potential targets:

- Protein-centric approach
- Pathway-centric approach

4.3.1 Protein-centric approach

After the binary comparisons, the proteins significantly altered in association with the silencing of ***all*** the genes were identified. Figure 33 shows the Venn diagram of the number of proteins significantly altered by siRNA.

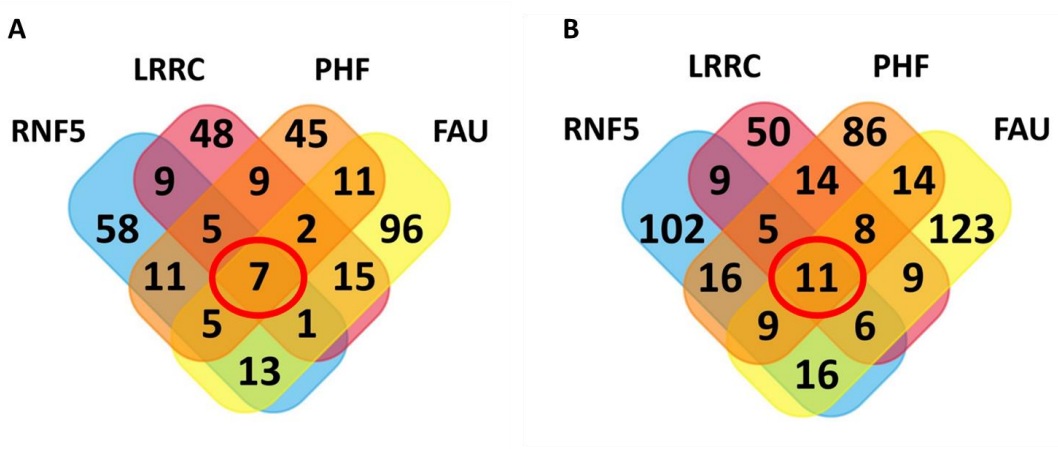


Figure 33: Venn diagram of the number of the proteins that were significantly dysregulated by the gene silencing. Panel A shows the number for the up-regulated proteins and Panel B shows the number of the down-regulated proteins. Red circles highlight the number of the proteins significantly altered in common among all the gene silencing experiments (7 and 11 for up- and down-regulated respectively).

Among the seven proteins significantly up-regulated by all the groups, there is CFTR, whose levels of expression are shown in Figure 34. This result shows that the ion library (Paragraph 3.1) is highly specific for CF research and that the gene maneuvers worked as expected and a significant CFTR rescue is observed in all the groups.

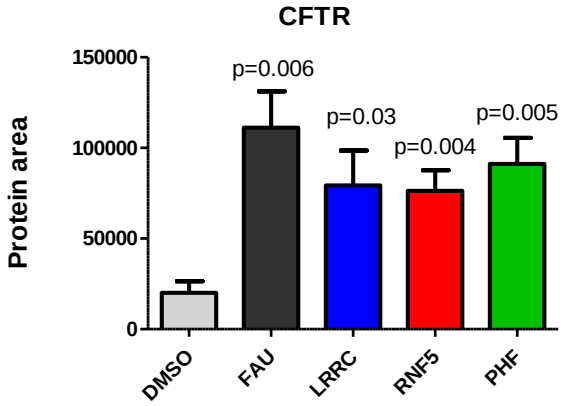


Figure 34: Histogram graph to compare the expression of CFTR in each group. The p-values derive from a t-test (un-paired, 2 tails t-test; N=4 for DMSO and N=5 for siRNAs) are shown each group in comparison to the control (DMSO).

Table 5 shows a description of all the seven proteins significantly up-regulated in association with all the siRNA.

Gene name	Protein name	Protein function
HSP90B1	**Endoplasmin**	Molecular chaperone that functions in the processing and transport of secreted proteins
CDV3	**Protein CDV3 homolog**	Role in cell proliferation
SCO1	**Protein SCO1 homolog, mitochondrial**	Copper metallochaperone essential for the maturation of cytochrome c oxidase subunit II
EHD1	**EH domain-containing protein 1**	ATP- and membrane-binding protein that acts in early endocytic membrane fusion and membrane trafficking of recycling endosomes
AATM	**Aspartate aminotransferase, mitochondrial**	Catalyzes the irreversible transamination of the L-tryptophan metabolite, L-kynurenine to form kynurenic acid (KA). Plays a key role in amino acid metabolism and in metabolite exchange between mitochondria and cytosol. Facilitates cellular uptake of long-chain free fatty acids
GTF2F1	**General transcription factor IIF subunit 1**	General transcription initiation factor that binds to RNA polymerase II, promoting the transcription elongation
CFTR	**Cystic fibrosis transmembrane conductance regulator**	Ion channel that, when mutated, causes CF

Table 5: List of the seven significantly up-regulated proteins: for each protein, described by both gene and protein name, a brief description of its function is shown.

From this list of proteins, four proteins were selected as possible direct effectors for CFTR functional rescue:

- **GTF2F1** (General transcription factor IIF subunit 1), which is a general transcription factor acting on several genes with functions not completely known [192].
- **CDV3** (Protein CDV3 homolog), which may have a role in cell proliferation [193].
- **EHD1** (EH domain-containing protein 1), able to bind both ATP and membrane, which acts in early endocytic membrane fusion and membrane trafficking of recycling endosomes [194].
- **HSP90B1** (Endoplasmin), a molecular chaperone directly involved in CFTR processing and folding [106].

Table 6 reports the proteins significantly down-regulated upon gene suppression of all the four proteins.

Gene name	Protein name	Function
CBS	**Cystathionine beta-synthase**	Involved in L-cystathionine synthesis; it can bind E3 ubiquitin ligase
PHGDH	**D-3-phosphoglycerate dehydrogenase**	Oxidoreductase involved in several pathways
MAPKP13	**Mitogen-activated protein kinase 13**	Serine-threonine kinase involved in the response to osmotic stress
CDK1	**Cyclin-dependent kinase 1**	Kinase involved in activation of MAPK cascade
TGM2	**Protein-glutamine gamma-glutamyltransferase 2**	Catalyzes the cross-linking of proteins and the conjugation of polyamines to proteins.
MAL2	**Protein MAL2**	Member of the machinery of polarized transport.
UHRF1	**E3 ubiquitin-protein ligase UHRF1**	Multidomain protein that acts as a key epigenetic regulator (by bridging DNA methylation and chromatin modification) and as E3 ubiquitin-protein ligase. It is still unclear how the two activities are related.
SMARCE1	**SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1**	Involved in transcriptional activation and repression of select genes by chromatin remodelling (alteration of DNA-nucleosome topology).
PRDX5	**Peroxiredoxin-5, mitochondrial**	Thiol-specific peroxidase that plays a role in cell protection against oxidative stress by detoxifying peroxides and as a sensor of hydrogen peroxide-mediated signalling events.
PCBP2	**Poly(rC)-binding protein 2**	Nucleic acid-binding protein. Negatively regulates cellular antiviral responses and is involved in proteasome-mediated protein catabolic process.
FN1	**Fibronectin**	Fibronectins bind cell surfaces and various compounds including chaperones

Table 6: List of the eleven proteins significantly down-regulated by all the gene silencing: for each protein, described by both gene and protein name, a brief description of its function is shown.

From the 11 **down-regulated** proteins in association with all the siRNA silencing, six proteins were selected as putative direct effectors for CFTR functional rescue:

- **SMARCE1** (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1), which is involved in transcriptional activation and repression of select genes by chromatin remodelling [195].
- **PCBP2** (Poly(rC)-binding protein 2), which can be involved in ubiquitin ligase activity [195].
- **CBS** (Cystathionine beta-synthase), which is an adapter of the E3 ubiquitin ligase [196].

- **MAL2** (Protein MAL2), which is a member of the machinery of polarized transport; required for the indirect transcytotic route [197].
- **UHRF1** (E3 ubiquitin-protein ligase UHRF1), a ubiquitin ligase known to be involved in DNA methylation and chromatin modification [198] but it can be involved also in CFTR degradation.
- **MAPK13** (Mitogen-activated protein kinase 13), a serine/threonine kinase which plays an important role in the cascades of cellular responses evoked by extracellular stimuli such as pro-inflammatory cytokines or physical stress. Its inhibition can alter airway mucus production [199].

From this experiment, a total of ten proteins were selected as putative direct effectors of CFTR functional rescue. In order to understand whether the selected proteins are involved in CFTR functional rescue or if their dysregulation is a side effect of the gene maneuvers, CFTR activity was measured after the gene suppression of the selected proteins, through the YFP-based assay (Paragraph 3.3.2). The suppression of proteins down-regulated in the first experiment should, in principle, favor CFTR, while suppression of the up-regulated proteins should play against CFTR rescue caused by FAU gene silencing.

For this reason, the expression of the ten selected proteins was suppressed either alone or in combination with FAU in CFBE41o- cells. Figure 35 shows the results of the YFP-based assay to evaluate CFTR activity in different conditions. To obtain a more robust outcome, 3 individual cell preparations, each with 5 individual biological replicates, were prepared for each condition.

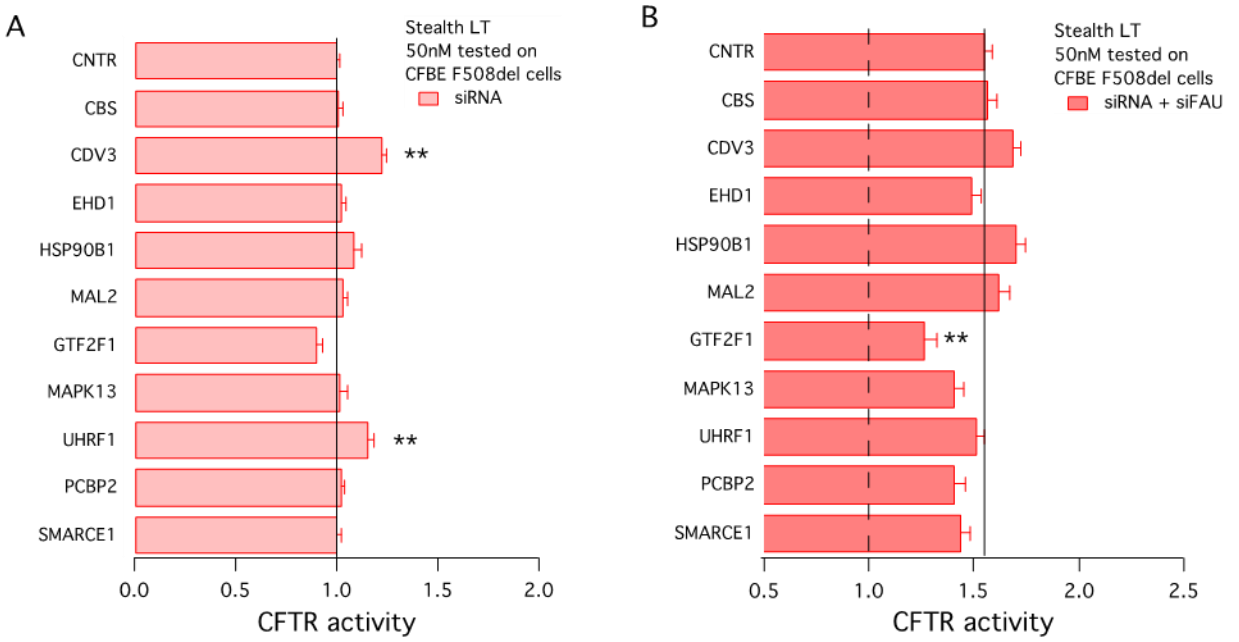


*Figure 35: CFTR rescue following silencing of the 10 selected proteins tested each independently (A) or in combination with FAU suppression (B). Three independent cell preparations, each with 5 biological replicates of CFBE41o- cells were used; genes silenced with pools of 3 siRNA, 50 nM final concentration. Statistical significance was assessed by parametric ANOVA, followed by Dunnett'st multiple post-hoc test. **=p<0.01 versus control (cells transfected with Non-Targeting siRNA Pool).*

From panel A, it is clear that two proteins (CDV3 and UHRF1), when silenced, trigger an increase in CFTR activity in comparison with the control. Since CDV3 was up-regulated by the first gene silencing, the increase in CFTR activity after its suppression shows that CDV3 negatively affects CFTR rescue and

processing. Panel B shows that the silencing of GTF2F1 protein triggers a significant down-regulation of CFTR activity when combined with the gene silencing of FAU.

To address this point, GTF2F1 silencing was performed in combination with either FAU or RNF5 silencing. As shown in Figure 36, FAU silencing is almost ineffective in rescuing CFTR when performed together with GTF2F1 silencing. On the other hand, CFTR rescue following RNF5 silencing is not impaired when GTF2F1 is absent.

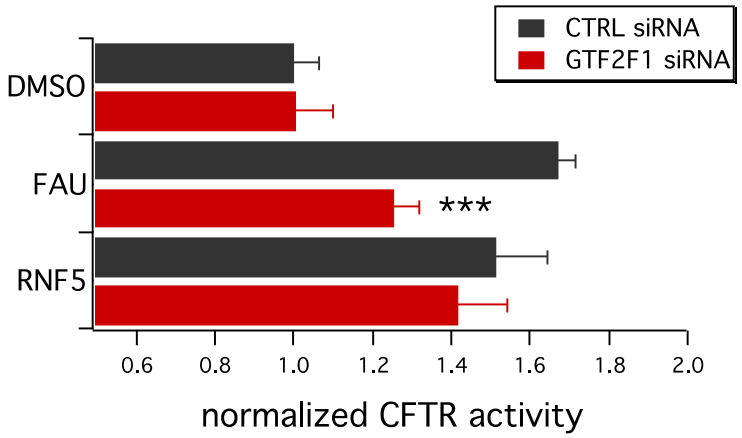


*Figure 36: CFTR rescue following GTF2F1 silencing together with either FAU or RNF5 silencing. Five independent cell preparations, each with 5 biological replicates, genes silenced with pools of 3 siRNA, 50 nM final concentration were tested. Statistical significance was assessed by parametric ANOVA, followed by Dunnett's multiple post-hoc test. **=p < 0.01 versus control (cells transfected with Non-Targeting siRNA Pool).*

These results suggested that CFTR rescue following FAU silencing could somehow be mediated by GTF2F1. These data could be very helpful in understanding the mechanism of action of FAU on CFTR rescue, which is still unknown. Further experiments are currently ongoing to specifically address this point.

4.3.2 Pathway-centric approach

In order to find other new putative targets for CF treatment, we then focused on the *biological pathways* – rather than proteins – altered in association with the silencing maneuvers. We investigated these pathways by performing gene enrichment search, by using the Reactome database (Paragraph 3.2.4). Reactome was interrogated with the dysregulated proteins from each of the groups and the resulting pathway identifiers were listed and filtered (only pathways identified by at least two peptides and showing the value of FDR <0.02 were retained). With these settings, four biological pathways significantly downregulated by all the gene silencing were found and are showed in Table 7.

Pathway identifier	Pathway Name	FDR
R-HSA-70614	Amino acid synthesis and interconversion (transamination)	0,19
R-HSA-75035	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B: Cdk1 complex	0,08
R-HSA-379716	Cytosolic tRNA aminoacylation	0,15
R-HSA-5628897	TP53 Regulates Metabolic Genes	0,10

Table 7: Biological pathways significantly downregulated following CFTR rescuing maneuvers. Each pathway is described with its pathway identifier, name and FDR (probability corrected for multiple comparisons) value, provided by Reactome databases [159].

Eleven proteins, known as being among the key regulators of these pathways, were selected for further investigation, by manually scrutinizing all the pathways data available. The list of the selected proteins is showed in

Table 8.

Gene name	Protein name	Function
YWHAZ	**14-3-3 protein zeta/delta**	Adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathways.
YWHAG	**14-3-3 protein gamma**	Adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathways.
GLS	**Glutaminase kidney isoform, mitochondrial**	Catalyzes the first reaction in the primary pathway for the renal catabolism of glutamine. Plays a role in maintaining acid-base homeostasis. Regulates the levels of the neurotransmitter glutamate.
PRDX2	**Peroxiredoxin-2**	Plays a role in cell protection against oxidative stress by detoxifying peroxides and as sensor of hydrogen peroxide-mediated signaling events. Might participate in the signaling cascades of growth factors and tumor necrosis factor-alpha.
PRDX5	**Peroxiredoxin-5**	Plays a role in cell protection against oxidative stress by detoxifying peroxides and as sensor of hydrogen peroxide-mediated signaling events.
LAMTOR1	**Ragulator complex protein LAMTOR1**	Involved in amino acid sensing and activation of mTORC1, a signaling complex promoting cell growth in response to growth factors, energy

		levels, and amino acids. Also required for late endosomes/lysosomes biogenesis. May be involved in cholesterol homeostasis.
LAMTOR2	**Ragulator complex protein LAMTOR2**	Involved in amino acid sensing and activation of mTORC1, a signaling complex promoting cell growth in response to growth factors, energy levels, and amino acids.
SERINC3	**Serine incorporator 3**	Restriction factor required to restrict infectivity of lentiviruses: acts by inhibiting an early step of viral infection. Impairs the penetration of the viral particle into the cytoplasm.
AIMP1	**Aminoacyl tRNA synthase complex-interacting multifunctional protein 1**	Non-catalytic component of the multisynthase complex. Stimulates the catalytic activity of cytoplasmic arginyl-tRNA synthase. Possesses inflammatory cytokine activity. Negatively regulates TGF-beta signaling. Involved in glucose homeostasis. Promotes dermal fibroblast proliferation. Regulates KDELR1-mediated retention of HSP90B1/gp96 in the endoplasmic reticulum. Plays a role in angiogenesis. Induces maturation of dendritic cells and monocyte cell adhesion. Modulates endothelial cell responses.
AIMP2	**Aminoacyl tRNA synthase complex-interacting multifunctional protein 2**	Required for assembly and stability of the aminoacyl-tRNA synthase complex. Mediates ubiquitination and degradation of FUBP1. Blocks MDM2-mediated ubiquitination and degradation of p53/TP53. Functions as a pro-apoptotic factor.
PPP2R5D	**Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform**	The B regulatory subunit might modulate substrate selectivity and catalytic activity, and also might direct the localization of the catalytic enzyme to a particular subcellular compartment.

Table 8: List of the eleven proteins, which are key regulators of the pathways significantly altered by all the four gene silencing. Each protein is described by gene name, protein name and a brief description of its function.

YFP-based assay (Paragraph 3.3.2) was performed to measure CFTR activity after the gene silencing of each of these proteins. The results of this assay are shown in Figure 37.

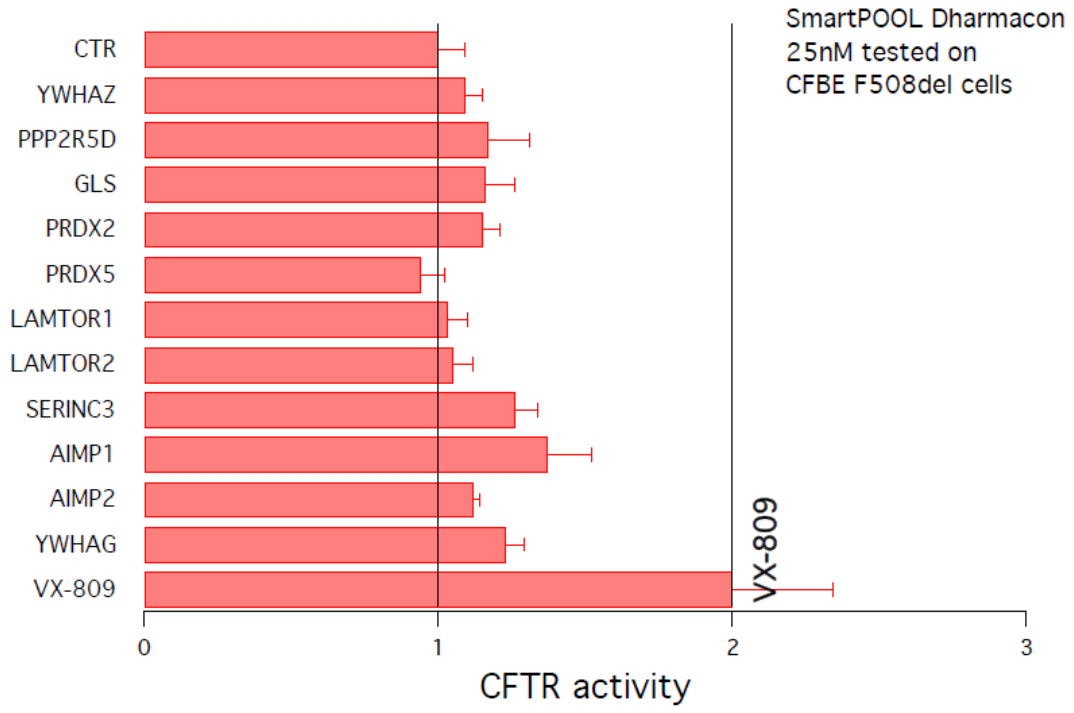


Figure 37: Preliminary experiment (one cell preparation with 5 biological replicates) showing CFTR rescue following silencing of 11 proteins selected with the pathway-centric approach.

From these results, two proteins (SERINC3 and AIMP1), when silenced, appear to trigger CFTR functional rescue.

The YFP-based assay was performed also to measure CFTR activity after the gene silencing of each of these proteins either independently or in combination with VX-809 treatment. The results of this assay are shown in Figure 38.

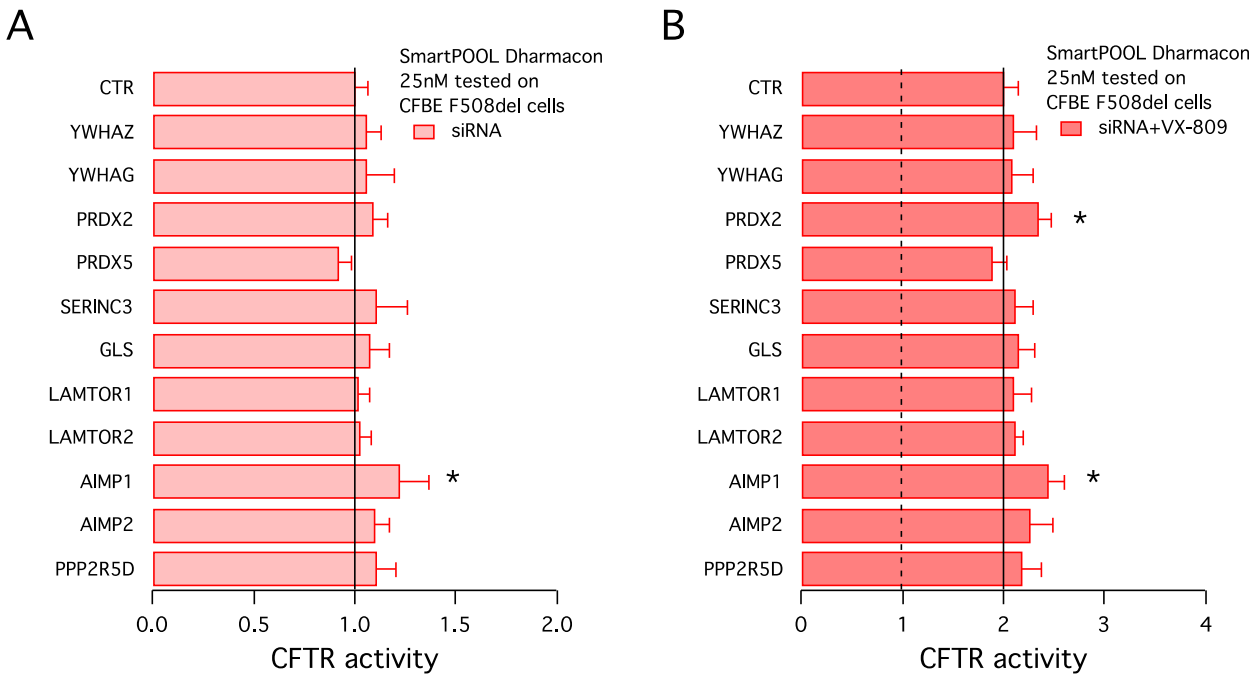


*Figure 38: Results of the YFP-based assay to measure CFTR rescue following silencing of 11 proteins selected with the pathway-centric approach either independently or in combination with VX-809 treatment (1 μM). These are results of four independent cell preparations with 5 biological replicates, genes silenced with pools of 4 siRNA, 100 nM final concentration. Black line corresponds to the corresponding control (activity on cells transfected with non targeting siRNA in A, and activity in cells after treatment with VX-809 1μM in B). Statistical significance was assessed by parametric ANOVA, followed by Dunnett's multiple post-hoc test. **=p < 0.01 versus control (cells transfected with Non-Targeting siRNA Pool).*

These results show that two proteins (AIMP1 and PRDX2) trigger an increase in CFTR activity (about 20%) in combination with VX-809 treatment. Interestingly, the silencing of PRDX5, a protein belonging to the same protein family of PRDX2, does not trigger CFTR rescue.

Since, in CF research, it is believed that the restoration of 5-10% of CFTR activity may be of substantial therapeutic value [200], these preliminary results are quite promising and thus worth of further investigations.

4.3.3 Selection of putative targets and treatment with modulators

The aim of this experiment was to identify new potential targets for CF. To reach this aim, we performed a proteomic profiling of CFBE41o- cells following the silencing of four proteins known to trigger CFTR rescue. We then focused on the identification of changes associated with the genetic maneuvers both at protein and pathway levels. After an extensive literature search (Paragraph 3.3.4), 27 proteins (listed in Table 9) were selected as new potential targets.

Target	Name	Pathway
LARS	Leucyl-tRNA synthetase	aminoacyl-tRNA synthetase
UHRF1	E3 ubiquitin-protein ligase UHRF1	/
HSP90B1	Endoplasmin	/
CDC25C	Dual specificity phosphatase Cdc25C	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex

CCNB1	G2/mitotic-specific cyclin-B1	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
YWHAG	14-3-3 protein gamma	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
YWHAZ	14-3-3 protein zeta	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
CDK1	Cyclin-dependent kinase 1	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
PRKAG1	5'-AMP-activated protein kinase subunit gamma-1	TP53 Regulates Metabolic Genes
PRKAA1	5'-AMP-activated protein kinase catalytic subunit alpha-1	TP53 Regulates Metabolic Genes
PRKAA2	5'-AMP-activated protein kinase catalytic subunit alpha-2	TP53 Regulates Metabolic Genes
PRKAB1	5'-AMP-activated protein kinase subunit beta-1	TP53 Regulates Metabolic Genes
YWHAG	14-3-3 protein gamma	TP53 Regulates Metabolic Genes
YWHAE	14-3-3 protein epsilon	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
SFN	14-3-3 protein sigma	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
CDC25C	Dual specificity phosphatase Cdc25C	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
CHEK2	Serine/threonine-protein kinase Chk2	Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
KARS	Lysyl-tRNA synthetase	aminoacyl-tRNA synthetase
LARS	Leucyl-tRNA synthetase	aminoacyl-tRNA synthetase
TP53	Cellular tumor antigen p53	TP53 Regulates Metabolic Genes
PRKAB2- PRKAG1-PRKAA2-PRKAA1-PRKAG3-PRKAG2-PRKAB1	PROTEIN COMPLEX GROUP: AMP-activated protein kinase, AMPK	TP53 Regulates Metabolic Genes
TNX	Thioredoxin	TP53 Regulates Metabolic Genes
MDM4-TP53-MDM2	PROTEIN-PROTEIN INTERACTION	TP53 Regulates Metabolic Genes
YWHAE	14-3-3 protein epsilon	TP53 Regulates Metabolic Genes
HSP90	Endoplasmin	/
UHRF1	E3 ubiquitin-protein ligase UHRF1	/
PRDX2	Peroxiredoxin-2	TP53 Regulates Metabolic Genes

Table 9: List of putative targets. For each of them, one or more modulators were tested of CFBE41o- cells.

As already described above (Paragraph 3.3.4), we then used several databases to find compounds able to chemically modulate these pathways. The selection of the chemical entities to test was mainly performed based on the following criteria:

a) **Biological activity**: the modulator must efficiently act on its target(s).

b) **LogP**: this parameter is the index of the lipophilicity of a molecule and it was considered in order to select modulators that were soluble in the medium of the CFBE41o- cells.
c) **Safety**: the modulators should have no reported adverse effects.
d) **Availability**: the modulators to be tested must be commercially available.

A total of 45 modulators for the targets reported in Table 10 were tested at concentrations ranging from 0.04 µM to 10 µM (based on their solubility and activity data) on CFBE41o- cells, and CFTR rescue was assessed with YFP-based assay.

N	Name	Target	Action
1	Carfilzomib	UHRF1	Inhibition
2	3-(2-Amino-6-quinazolinyl)-4-methyl-1-[4-(1,3-oxazol-2-yl)phenyl]-2(1H)-pyridinone	PRKAB1	Inhibition
3	CCT244747	PRKAB2-PRKAG1-PRKAA2-PRKAA1-PRKAG3-PRKAG2-PRKAB1	Inhibition
4	Withaferin A	HSP90	Inhibition
5	Pifithrin-µ	TP53	Inhibition
6	Brotezomib	UHRF1	Inhibition
7	H-89 | H-89 Chloride	CHEK2	Inhibition
8	Seliciclib	CCNB1	Inhibition
9	(-)-Gambogic acid	TNX	Inhibition
10	1-(1H-Indol-3-ylmethyl)proline	CDC25C	Inhibition
11	Conoidin A	PRDX2	Inhibition
12	4-(2,6-DIMETHYL-4-HYDROXYPHENYLAZO)BENZENESULFONIC ACID	TP53	Inhibition
13	2-[(E)-{[4-(4-Methylphenyl)-6-phenyl-2-pyrimidinyl]hydrazono}methyl]-4-nitrophenol	SFN	Inhibition
14	7-Hydroxy-3-(4-hydroxyphenyl)-4-methyl-2H-chromen-2-one	CDC25C	Inhibition
15	5-{[1-(4-Chlorophenyl)-2,5-diphenyl-1H-pyrrol-3-yl]methylene}-2-thioxodihydro-4,6(1H,5H)-pyrimidinedione	MDM4-TP53-MDM2	Inhibition
16	(5E)-5-{[1-(4-Chlorophenyl)-2,5-diphenyl-1H-pyrrol-3-yl]methylene}-1-phenyl-2-thioxodihydro-4,6(1H,5H)-pyrimidinedione	MDM4-TP53-MDM2	Inhibition

17	2-({[2-(4-Methoxyphenyl)-5-methyl-1,3-oxazol-4-yl]methyl}sulfinyl)-N-(2-pyridinylmethyl)acetamide	YWHAG	Inhibition
18	Alpha-Amyrin	YWHAE	Inhibition
19	NSC-114945 | Oleanolic Acid Glutaryl Hemi Ester	YWHAE	Inhibition
20	(2S)-4-Methyl-1,2-pentanediamine	LARS	Inhibition
21	PF-3758309	PRKAA1	Inhibition
22	8-[(3,5-dichlorophenyl)sulfanyl]-9-{3-[(propan-2-yl)amino]propyl}-9H-purin-6-amine	HSP90B1	Inhibition
23	TG-02; (16E)-14-Methyl-20-oxa-5,7,14,27-tetraazatetracyclo[19.3.1.12,6.18,12]heptacosa-1(25),2(27),3,5,8(26),9,11,16,21,23-decaen	CDK1	Inhibition
24	(5Z)-5-[1-(Hydroxyamino)ethylidene]-2,4(3H,5H)-pyrimidinedione	UHRF1	Inhibition
25	QUINOLYLUREA	CDC25C	Inhibition
26	2-(4-Morpholinyl)-4H-pyrimido[2,1-a]isoquinolin-4-one	KARS	Inhibition
27	3-(3-Aminopropoxy)-N-[2-({3-[(4-chloro-2-hydroxyphenyl)carbamoyl]-2-naphthyl}oxy)ethyl]-2-naphthamide hydrochloride (1:1)	TP53	Inhibition
28	2-{3-[(4,6-dioxo-2-sulfanylidene-1,3-diazinan-5-ylidene)methyl]-2,5-dimethyl-1H-pyrrol-1-yl}-4,5,6,7-tetrahydro-1-benzothiophene-3-carbonitrile	MDM4-TP53-MDM2	Inhibition
29	(5Z)-5-(4-Propoxybenzylidene)-1,3-thiazolidine-2,4-dione	KARS	Inhibition
30	2-[(1,5-Dimethyl-3-oxo-2-phenyl-2,3-dihydro-1H-pyrazol-4-yl)carbamoyl]terephthalic acid	SFN	Inhibition
31	clofibrate	TP53	Inhibition
32	7-Azaindole | 1H-Pyrrolo[2,3-B]Pyridine	CHEK2	Inhibition
33	L-(+)-Leucinol	LARS	Inhibition
34	5-Acetyluracil	UHRF1	Inhibition
35	Methyl mesylate	TP53	Inhibition
36	Azacitidine	UHRF1	Inhibition
37	(4S)-4-(2-methylpropyl)-1,3-oxazolidin-2-one	LARS	Inhibition
38	(3S)-5-methylhexane-1,3-diamine	LARS	Inhibition
39	2-(2-methylpropyl)-1H-imidazole	LARS	Inhibition
40	(3S)-3-Amino-5-methyl-1-hexanol	LARS	Inhibition
41	2-(3,7-Dihydroxydibenzo[b,f][1,4]oxazepin-11-yl)benzoic acid	YWHAG	Inhibition
42	2,4-Dimethyl-2,3-dihydrothieno[3,2-c]quinoline-8-carboxylic acid	YWHAG	Inhibition
43	Novobiocin	HSP90	Inhibition

44	Sodium Vanadate | Vanadate	CDC25C	Inhibition
45	4-[(E)-{4-Formyl-5-hydroxy-6-methyl-3-[(phosphonooxy)methyl]-2-pyridinyl}diazenyl]benzoic acid	YWHAZ	Inhibition

Table 10: List of modulators of putative targets tested on CFBE41o- cells.

The corresponding results of the YFP-based assay to measure CFTR rescue following the treatment with the modulators listed in Table 10 are showed in Figure 39.

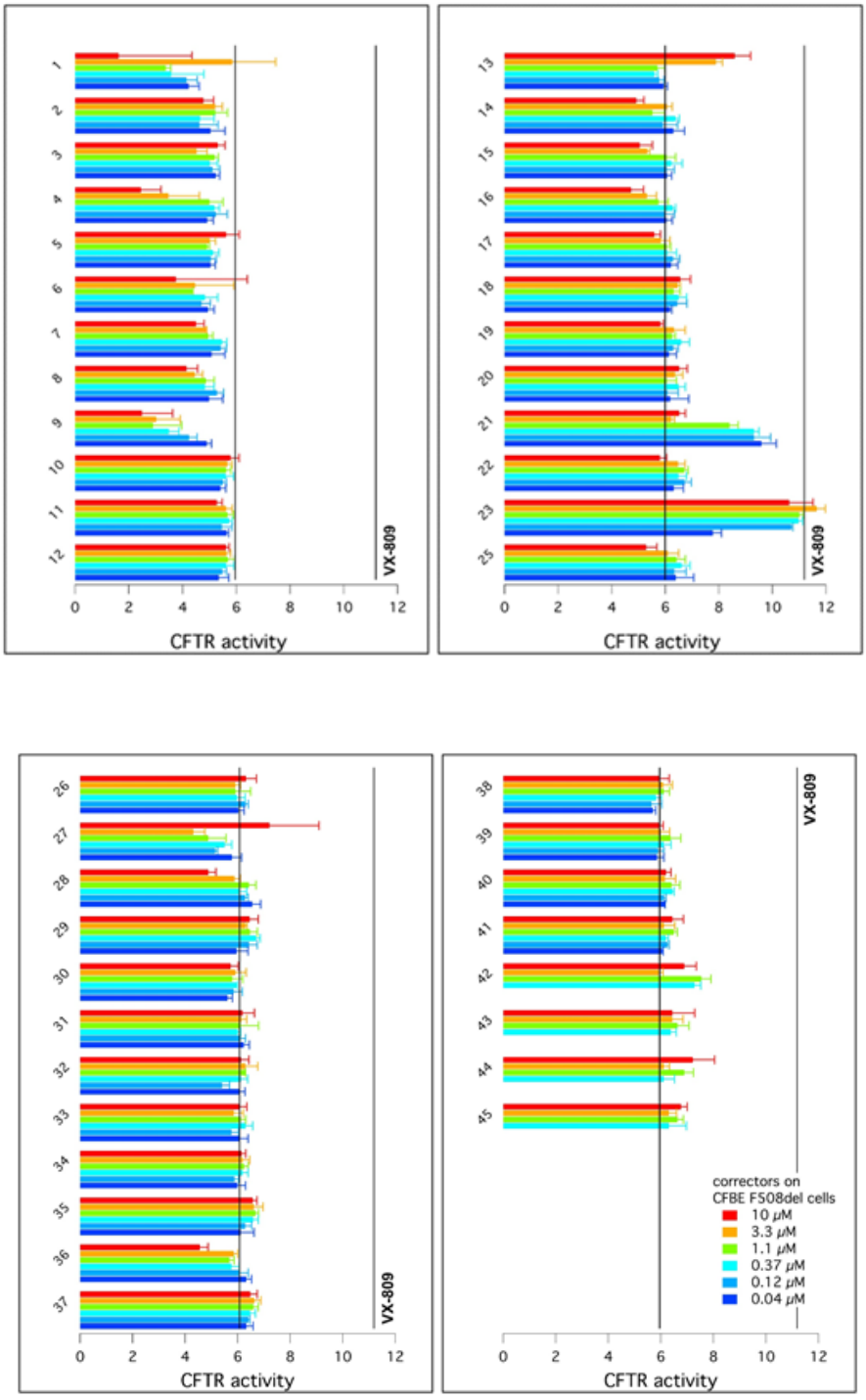


Figure 39: YFP-based assay to measure CFTR activity following the treatment with different modulators of putative targets on CFBE41o- cells.

Figure 39 shows that only the compound **23**, trigger a CFTR rescue that is comparable with the rescue of CFTR following VX-809 treatment. For this reason, only the compound 23 was tested on the BE primary cells derived from a patient homozygous for F508del-CFTR either alone or in combination with the corrector VX-445. The corresponding results of the short-circuit current recordings in Ussing chamber following the treatments are showed in Figure 40.

For this experiment, bronchial epithelia differentiated under air-liquid condition were treated for 24 hours by adding the compounds to the basolateral medium. The following day, the epithelia were mounted in a perfusion chamber to measure short-circuit current. After inhibition of the epithelial sodium channel (ENaC) with amiloride (10 µM), cells were stimulated with the membrane permeable cAMP analogue CPT-cAMP (100 µM) plus VX-770 (1 µM) to maximally activate F508del-CFTR. Then, we inhibited CFTR channel by adding the CFTR inhibitor-172 (10 µM). The amplitude of the current drop caused by inh-172 was taken as an estimate of total CFTR activity in each epithelium.

In epithelia treated for 24 hours with vehicle alone (DMSO) the total CFTR activity is very low (Figure 40). Treatment with compound **23** caused no increase in CFTR-mediated current (Figure 40). Interestingly, epithelia that were treated with VX-445 for 24h displayed remarkably augmented chloride secretion, with a 4-to-5-fold increase in CFTR total current (Figure 40). However, no further increase was observed when epithelia were treated with VX-445 and compound **23** (Figure 40).

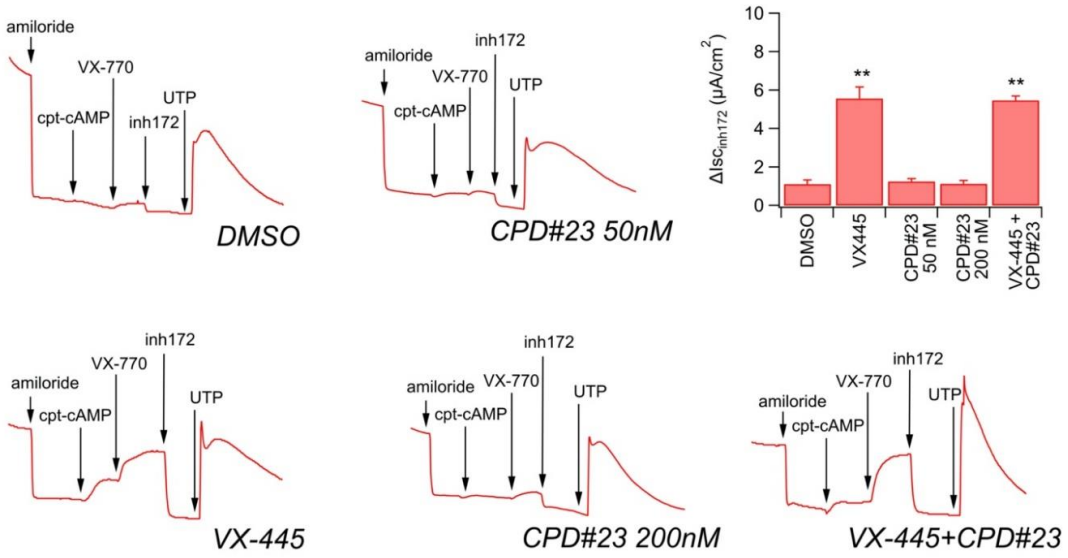


*Figure 40: Representative traces and summary of results obtained from short-circuit current recordings on F508del/F508del bronchial epithelial cells. Data reported as mean ± SD of the amplitude of the current blocked by inh-172 (10 µM) in epithelia treated with the indicated compounds. *P < 0.05 and **P < 0.01 versus control [analysis of variance (ANOVA) with Dunnett's post hoc test].*

Further experiments on other compounds selected are still on going. Compound **21** will also soon be tested on primary cells.

4.4 Lipidomics profiling of CFBE41o- after incubation with drugs

The importance of membrane lipids in the physiology and trafficking of CFTR was recently investigated: cholesterol and ceramides have a role in the formation of CFTR clusters at the plasma membrane [201], [202]; phosphatidylserines help in stabilizing CFTR at the plasma membrane [203]. Another very recent study shows the significant role of cholesterol in CFTR activity [115]. Other important players in the stabilization of CFTR at the PM are the flippases [204], enzymes that regulate the phospholipids movement across the membranes of the cells.

Moreover, recent studies points at a very significant role of the cell lipidome in the mechanism of action of CF drugs:

- A positive impact of pharmacological modulation on the sphingolipids signaling pathway in CF pathology has been proposed [205], [206].
- On the other hand, a negative effect on CFTR half-life at the plasma membrane has been associated with chronic co-administration of the potentiator ivacaftor (VX-770) and the corrector lumacaftor (VX-809) [207], [208].

Although all these studies gave an important contribution in elucidating the role of some lipids in CF, an untargeted lipidomics profile of the bronchial epithelial cells has not previously been performed. The aim of this experiment was to investigate the lipidome profile of the F508del-CFTR expressing cells following successful CFTR pharmacological rescue strategies. We hypothesized that the pharmacological rescue of mutant CFTR could be associated with changes in lipidome profile of CFBE41o- cells. All the data generated by the lipidomics experiment on the CFBE41o- (Paragraph 3.4) were analyzed by multivariate data analysis tools like PCA and PLS-DA to identify changes in the global lipidomic profiles.

For this experiment, several groups were analyzed:

- ***The treated groups***: cells were treated with different CF drugs (the corrector VX-809, the potentiator VX-770, the combination of VX-809 and VX-770 and the triple combination of VX-661/VX-445/VX-770).
- ***The control group***: cells were treated with DMSO.
- ***The QCs***: quality control samples resulting from the pooling of all the samples.
- ***The blank groups***: both procedural blanks and acquisition blanks for LC-MS/MS runs.

The PCA score plot revealed not only a clear separation between the groups (the QCs, the treated samples and the controls) and blank runs (Figure 41) but also differences between the five experimental groups, especially over PC1 *vs*. PC3 and PC 2 *vs*. PC3 (Figure 42).

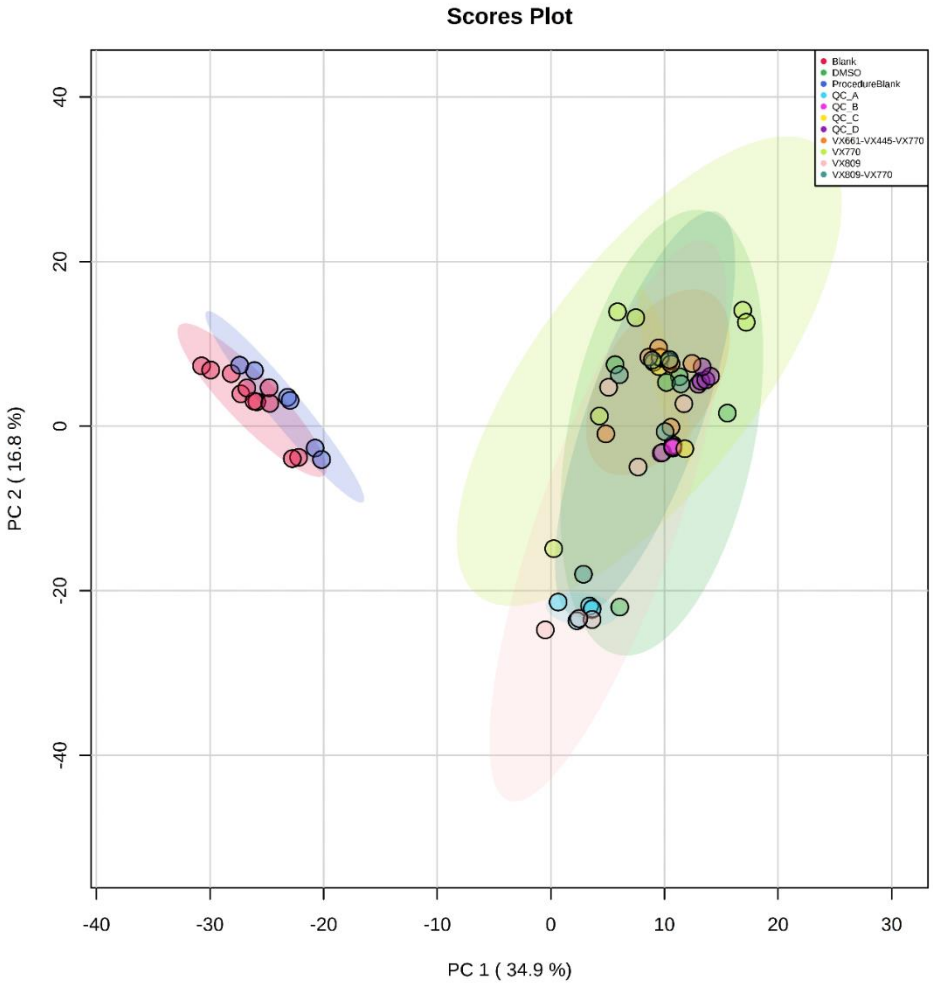


Figure 41: PCA score plot of all the features observed from all the experimental groups. Picture taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

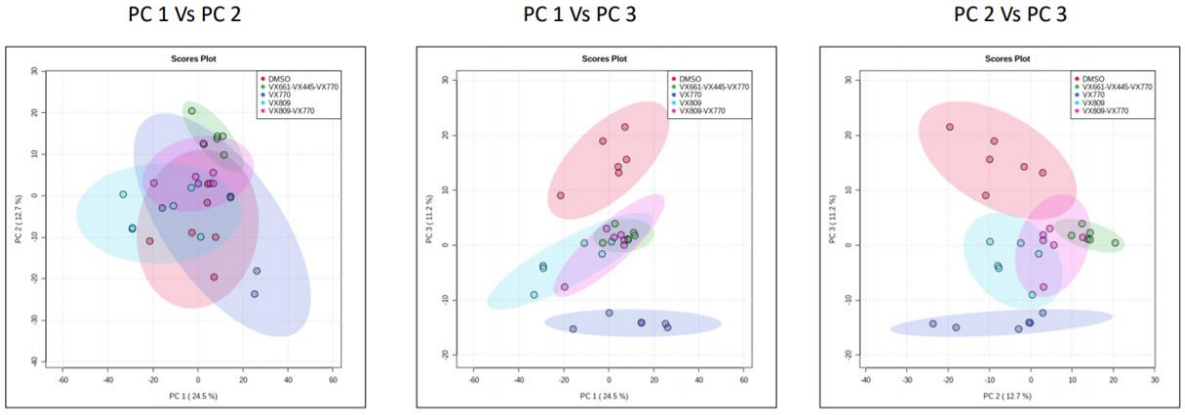


Figure 42: PCA score plots of all the features observed from all the experimental groups, different PCs were selected and indicated in the plots. Pictures taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

Another PCA score plot revealed that the group treated with the combination VX-809-VX-770 is well-separated from the groups treated with the molecules separately (shown in Figure 43).

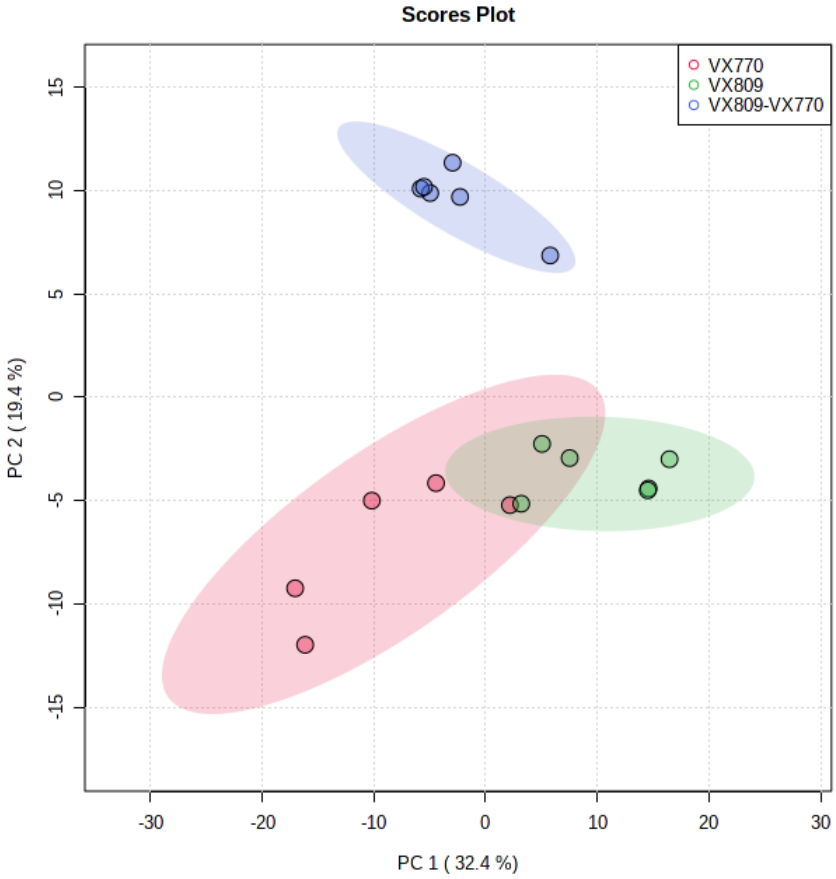


Figure 43: PCA score plot of all the features observed in the groups treated with either VX-809 and VX-770 alone or in combination.

The separation of all the groups was revealed by the PLS-DA plot, shown in Figure 44.

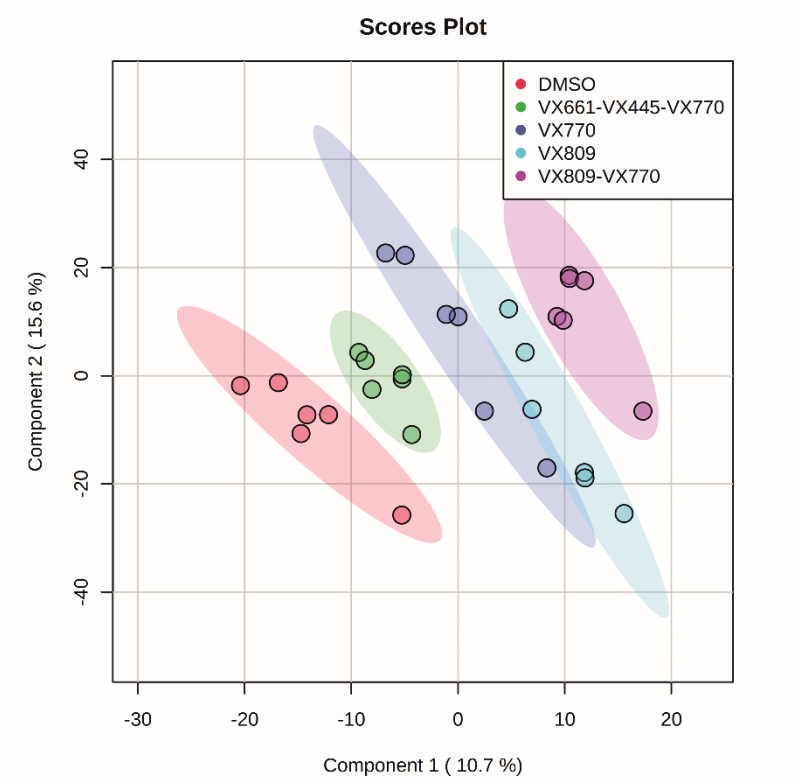


Figure 44: PLS-DA score plot of the lipidomics analysis of CFBE41o- after treatment with drugs. DSMO group represents the control, where the cells were treated only with the vehicle. Blank and QC groups are omitted for clarity. Picture taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

The leave one out cross-validation of PLS-DA analysis (Figure 45), demonstrated that the PLS-DA model for this experiment is reliable, and the model is not significantly overfitting the data.

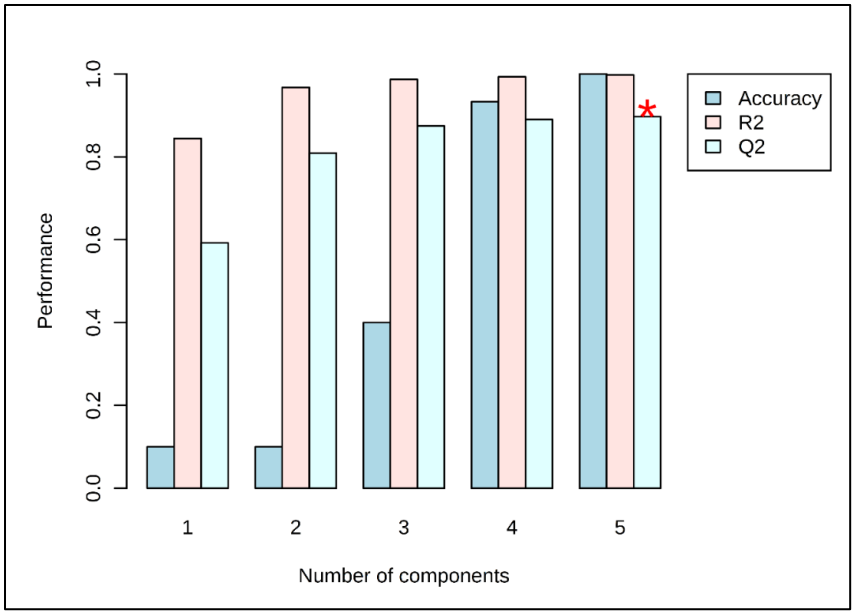


Figure 45: Results of leave-one-out cross validation of the PLS-DA model. Positive Q2 values (0.6 and 0.8 for 1 and 2 principal component respectively) indicate that the model is not overfitted.

The supervised analysis, accounting for 26% of the total PC variability observed in the dataset, showed that:

- ✓ The differences among the experimental groups exceeded the observed variability among the different biological replicates, represented by the spreading of the dots within the same group.
- ✓ Each treatment caused peculiar changes into the cell lipidome, resulting in a clear separation of all the groups. Based on VIP scores, 75 features (provisionally identified with their RTs and *m/z* values) were selected as those contributing the most to the groups' separation. Based on these signals, a clusterization analysis was performed and the resulting heatmap (Figure 46) shows that the samples naturally cluster into the five experimental groups. Using the LIPIDMAPS database [178], a total of 48 features (out of 75) were positively assigned to a lipid ID.

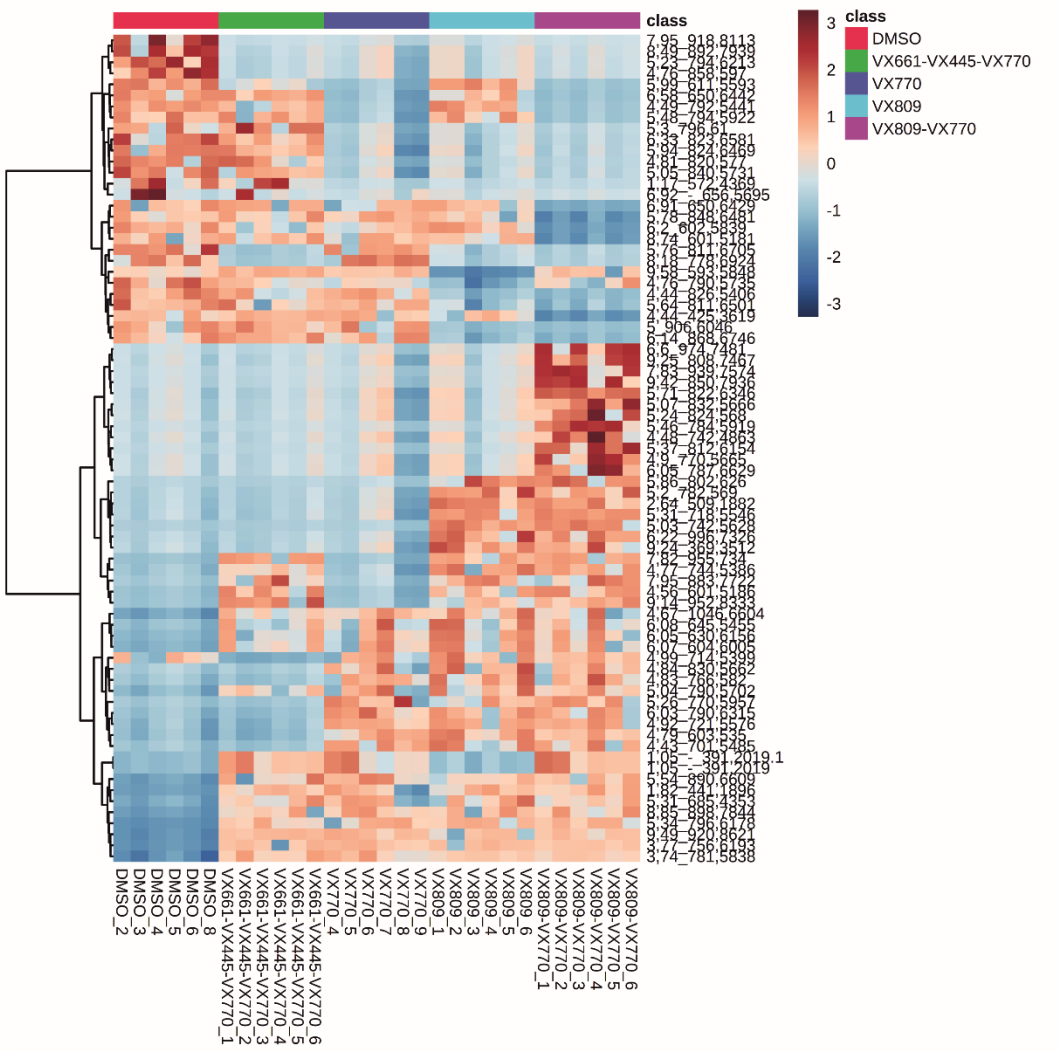


Figure 46: Heatmap resulting from the clustering analysis of the CFBE41o- lipidomics dataset. This heatmap was created using the top 75 PLS-DA VIP features. Picture taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

Moreover, a pattern correlation analysis (Figure 47) revealed the presence of two major clusters of activated lipid pathways: the glycerophospholipid and the sphingolipid metabolisms. These results indicate that the treatment with CF drugs alters two of the most important metabolic pathways for lipids.

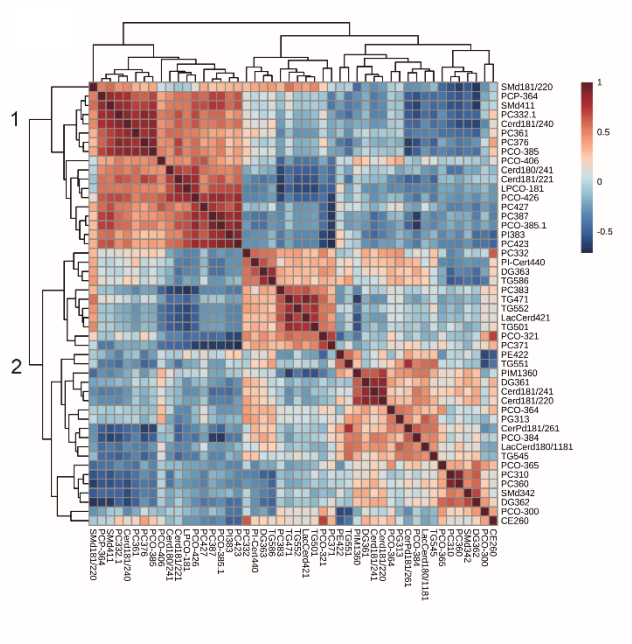


Figure 47: Patterns of correlation among the 48 lipids. Two main clusters of lipids were observed (labelled 1 and 2) Picture taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

Besides the multivariate data analysis, where the results of a simultaneous comparison of all the tested groups were analyzed all at once, binary comparisons were also performed to detect distinctive changes associated with each treatment.

4.4.1 Role of VX-770

The Venn diagram in Figure 48 highlights the overlap of the lipids significantly altered in all the conditions involving the use of VX-770.

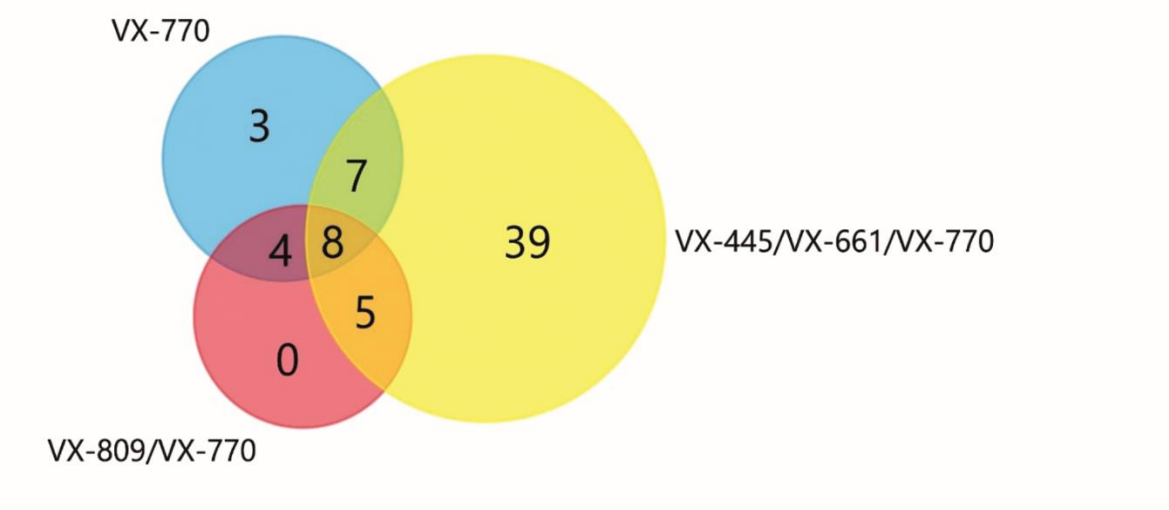
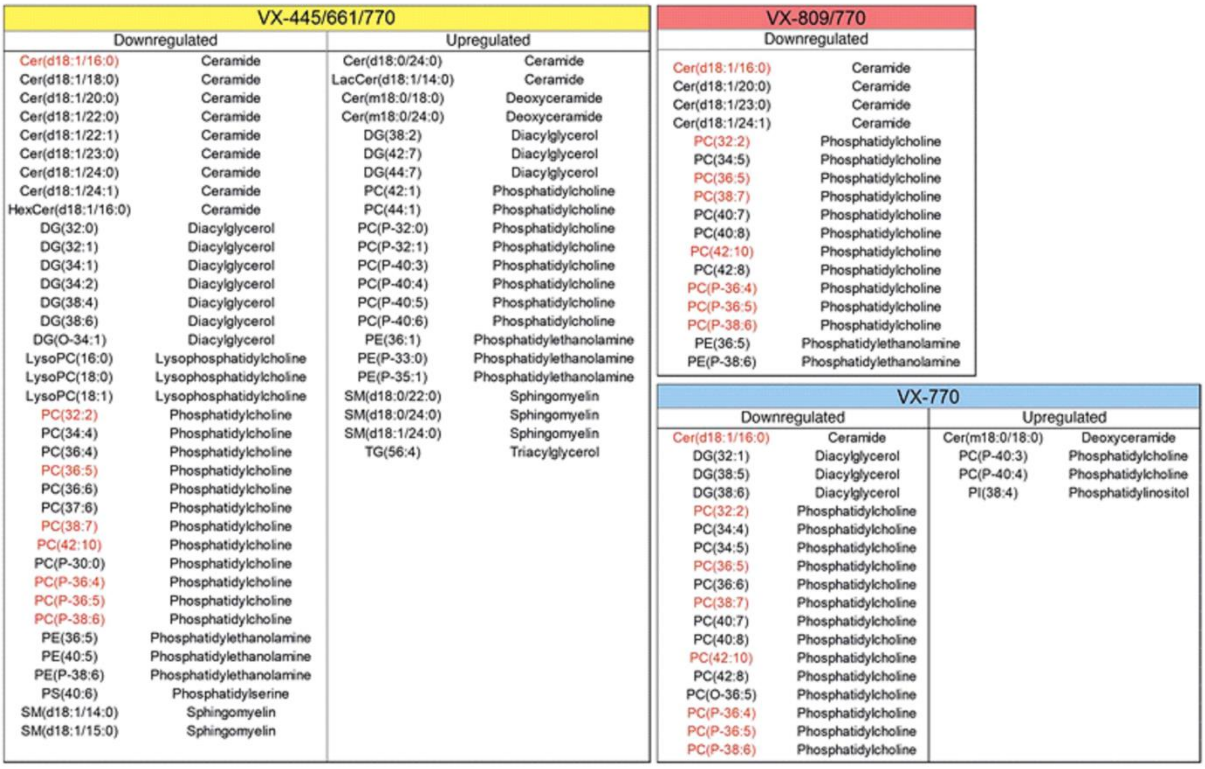


Figure 48: Venn diagram reporting significantly altered lipids (compared to control, FDR 0.05, FC>1.5) from the treatments involving VX-770. Picture taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

The treatment with Trikafta (VX-445/VX-661/VX-770) has the biggest impact on the lipidome profile of CFBE41o- cells, with 39 lipids altered only by this treatment. At the intersection of all the groups, 8 lipids are altered (all down-regulated compared to the control groups) by all the treatments involving VX-770, and thus appear to be specifically associated with this drug.



VX-445/661/770

Downregulated		Upregulated	
Cer(d18:1/16:0)	Ceramide	Cer(d18:0/24:0)	Ceramide
Cer(d18:1/18:0)	Ceramide	LacCer(d18:1/14:0)	Ceramide
Cer(d18:1/20:0)	Ceramide	Cer(m18:0/18:0)	Deoxyceramide
Cer(d18:1/22:0)	Ceramide	Cer(m18:0/24:0)	Deoxyceramide
Cer(d18:1/22:1)	Ceramide	DG(38:2)	Diacylglycerol
Cer(d18:1/23:0)	Ceramide	DG(42:7)	Diacylglycerol
Cer(d18:1/24:0)	Ceramide	DG(44:7)	Diacylglycerol
Cer(d18:1/24:1)	Ceramide	PC(42:1)	Phosphatidylcholine
HexCer(d18:1/16:0)	Ceramide	PC(44:1)	Phosphatidylcholine
DG(32:0)	Diacylglycerol	PC(P-32:0)	Phosphatidylcholine
DG(32:1)	Diacylglycerol	PC(P-32:1)	Phosphatidylcholine
DG(34:1)	Diacylglycerol	PC(P-40:3)	Phosphatidylcholine
DG(34:2)	Diacylglycerol	PC(P-40:4)	Phosphatidylcholine
DG(38:4)	Diacylglycerol	PC(P-40:5)	Phosphatidylcholine
DG(38:6)	Diacylglycerol	PC(P-40:6)	Phosphatidylcholine
DG(O-34:1)	Diacylglycerol	PE(36:1)	Phosphatidylethanolamine
LysoPC(16:0)	Lysophosphatidylcholine	PE(P-33:0)	Phosphatidylethanolamine
LysoPC(18:0)	Lysophosphatidylcholine	PE(P-35:1)	Phosphatidylethanolamine
LysoPC(18:1)	Lysophosphatidylcholine	SM(d18:0/22:0)	Sphingomyelin
PC(32:2)	Phosphatidylcholine	SM(d18:0/24:0)	Sphingomyelin
PC(34:4)	Phosphatidylcholine	SM(d18:1/24:0)	Sphingomyelin
PC(36:4)	Phosphatidylcholine	TG(56:4)	Triacylglycerol
PC(36:5)	Phosphatidylcholine		
PC(36:6)	Phosphatidylcholine		
PC(37:6)	Phosphatidylcholine		
PC(38:7)	Phosphatidylcholine		
PC(42:10)	Phosphatidylcholine		
PC(P-30:0)	Phosphatidylcholine		
PC(P-36:4)	Phosphatidylcholine		
PC(P-36:5)	Phosphatidylcholine		
PC(P-38:6)	Phosphatidylcholine		
PE(36:5)	Phosphatidylethanolamine		
PE(40:5)	Phosphatidylethanolamine		
PE(P-38:6)	Phosphatidylethanolamine		
PS(40:6)	Phosphatidylserine		
SM(d18:1/14:0)	Sphingomyelin		
SM(d18:1/15:0)	Sphingomyelin		

VX-809/770

Downregulated	
Cer(d18:1/16:0)	Ceramide
Cer(d18:1/20:0)	Ceramide
Cer(d18:1/23:0)	Ceramide
Cer(d18:1/24:1)	Ceramide
PC(32:2)	Phosphatidylcholine
PC(34:5)	Phosphatidylcholine
PC(36:5)	Phosphatidylcholine
PC(38:7)	Phosphatidylcholine
PC(40:7)	Phosphatidylcholine
PC(40:8)	Phosphatidylcholine
PC(42:10)	Phosphatidylcholine
PC(42:8)	Phosphatidylcholine
PC(P-36:4)	Phosphatidylcholine
PC(P-36:5)	Phosphatidylcholine
PC(P-38:6)	Phosphatidylcholine
PE(36:5)	Phosphatidylethanolamine
PE(P-38:6)	Phosphatidylethanolamine

VX-770

Downregulated		Upregulated	
Cer(d18:1/16:0)	Ceramide	Cer(m18:0/18:0)	Deoxyceramide
DG(32:1)	Diacylglycerol	PC(P-40:3)	Phosphatidylcholine
DG(38:5)	Diacylglycerol	PC(P-40:4)	Phosphatidylcholine
DG(38:6)	Diacylglycerol	PI(38:4)	Phosphatidylinositol
PC(32:2)	Phosphatidylcholine		
PC(34:4)	Phosphatidylcholine		
PC(34:5)	Phosphatidylcholine		
PC(36:5)	Phosphatidylcholine		
PC(36:6)	Phosphatidylcholine		
PC(38:7)	Phosphatidylcholine		
PC(40:7)	Phosphatidylcholine		
PC(40:8)	Phosphatidylcholine		
PC(42:10)	Phosphatidylcholine		
PC(42:8)	Phosphatidylcholine		
PC(O-36:5)	Phosphatidylcholine		
PC(P-36:4)	Phosphatidylcholine		
PC(P-36:5)	Phosphatidylcholine		
PC(P-38:6)	Phosphatidylcholine		

Table 11: Summary of the lipids dysregulated by each treatment involving VX-770. The red lipids are in common among all the treatments involving VX-770. Figure adapted from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

As shown in Table 11, among the 8 down-regulated lipids, the phosphatidylcholines, lipids which are among the major components of cell membranes [209], are present. These molecules have a huge impact on the mechanical and biophysical properties of the membranes [210]. It was already demonstrated that changes in the metabolism of this class of lipids are linked to the altered membrane recycling in CF [211]. Moreover, these lipids are one of the major components of BAL (Broncho-Alveolar Lavage) fluid in CF patients [212]. These results are consistent with other studies already conducted by Bear's group, demonstrating that VX-770 increases membrane fluidity, by interacting with membrane lipid structures [213]. Another study [214] demonstrated that VX-770 restores the secretion of the phosphatidylcholine lipids, which is mediated by the ABCB4 membrane transporter. This increased secretion can cause the decrease of the individual phosphatidylcholine lipids detected in this lipid experiment.

The observed downregulation of ceramide is also intriguing, since ceramides are known to accumulate in CF epithelial cells, causing inflammation and cell death [215]. Moreover, ceramides were associated with cell membrane proprieties [216] and with the lipid rafts formation [216].

4.4.2 Focus on the effect of the triple combination (Kaftrio)

Since Kaftrio is the latest drug approved for CF therapy [49], its global effects on cell lipidome have not been investigated yet. As shown in Table 11, Kaftrio has the major effect on CFBE41o- cells lipidome, causing the upregulation of 22 lipids and the downregulation of 37 lipids. Figure 49 shows that the control and the treated (with Kaftrio) group are already clearly separated on the PCA score plot.

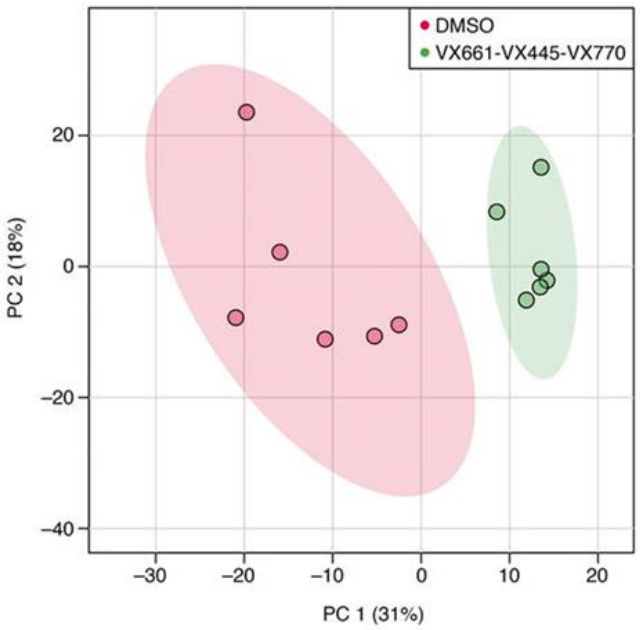


Figure 49: PCA score plot of the control and triple combination (VX-661/VX-445/VX-770, Trikafta) comparison showing a clear separation of the groups in the PC1 and PC2. Figure adapted from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta [149].

From this untargeted approach, ceramides resulted the most altered lipids (downregulated) by this treatment. To confirm these results, a targeted analysis (Paragraph 3.4.4) was performed to specifically measure the levels of six ceramides in all the groups. These six lipids, [Cer(d18:1/16:0), Cer(d18:1/18:0), Cer(d18:1/20:0), Cer(d18:1/22:0), Cer(d18:1/24:1), Cer(d18:1/24:0)], are among the most abundant ceramides in mammalian tissues [217].

Figure 50 shows the results of the targeted analysis: only the treatment with Kaftrio is able to significantly downregulate the ceramides in all the groups, although a trend of downregulation of this class of lipids can be observed even with other treatments (VX-770 alone and the combination VX-770/VX-809).

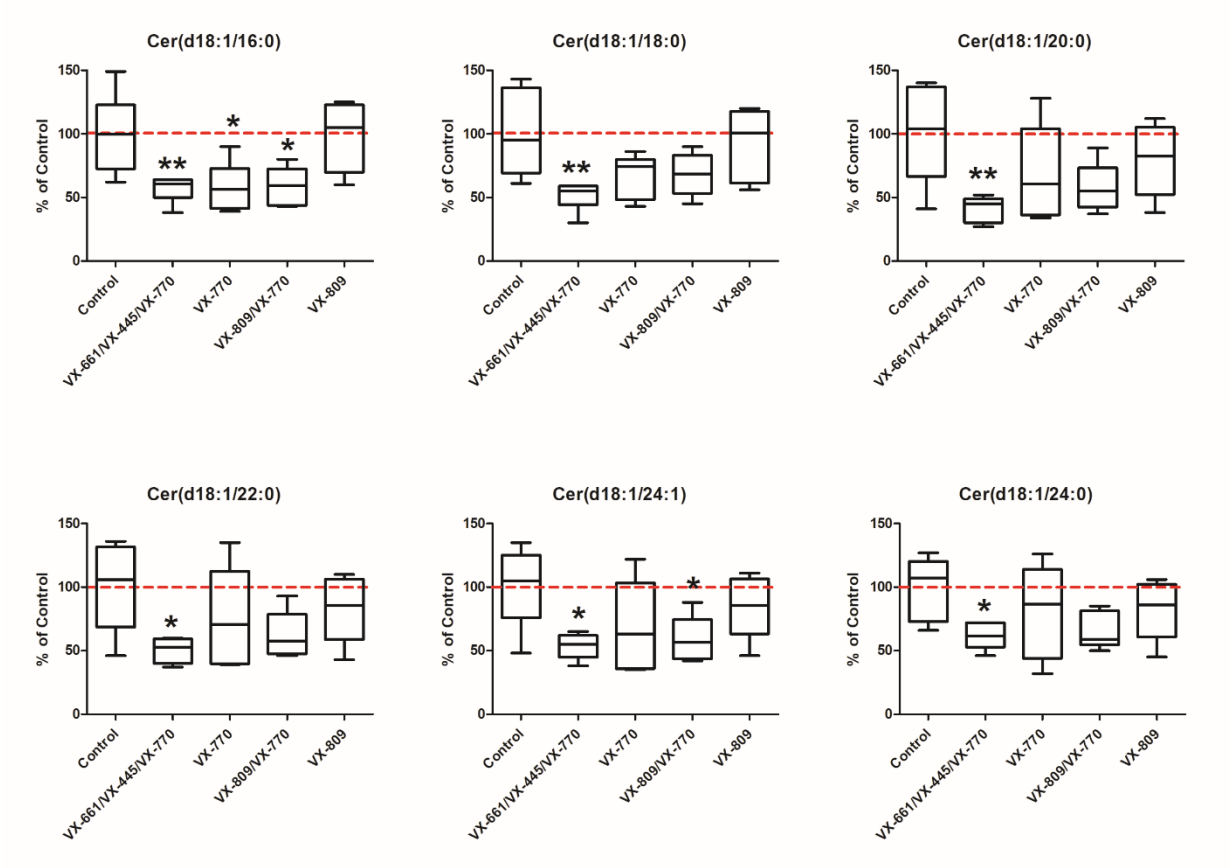


*Figure 50: Results of the targeted quantification of six ceramides in all the experimental groups. The dotted red line represents the average level of the control group. Box-and-whiskers plots represent the median value (inner bar), the first and third quartile (box), and the minimum and maximum value (outer bars). Data represent mean ± SEM, n = 5. *P < 0.05, **P < 0.01 compared with control, 1-way ANOVA test with Dunnett'st's post hoc test. Figure adapted from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta* [149].

Ceramides are known to be involved in the induction of apoptosis [218]; moreover, it was demonstrated that the treatment with an inhibitor of ceramide biosynthesis (XM462) promoted cell survival in gastric carcinoma HCG27 cell line [219]. Given the known role of ceramides, the biological relevance of these results was further investigated through the evaluation of the response of the CFBE41o- cells to pro-apoptotic stimuli after the treatment with Kaftrio (Paragraph 3.4.5). The hypothesis was that, by lowering the ceramide levels by pharmacological treatment, the cells should have shown a decrease in susceptibility to apoptosis.

CFBE41o- cells expressing F508del-CFTR were then exposed to different concentrations (1–10 µM) of etoposide, a topoisomerase II inhibitor able to induce apoptosis in bronchial cells (Paragraph 3.4.5). The following day, the number of living cells was evaluated for each condition (Figure 51). The results showed that none of the pharmacological treatments that were investigated affects the number of living cells under resting condition. Nevertheless, when a pro-apoptotic stimulus (etoposide) is given, the number of viable cells decreases following a dose-dependent way. However, when etoposide is added to the cells in the presence of the Kaftrio treatment, the number of viable cells is consistently higher in comparison to their control (cells treated only with etoposide) as showed in Figure 51, Panel B.

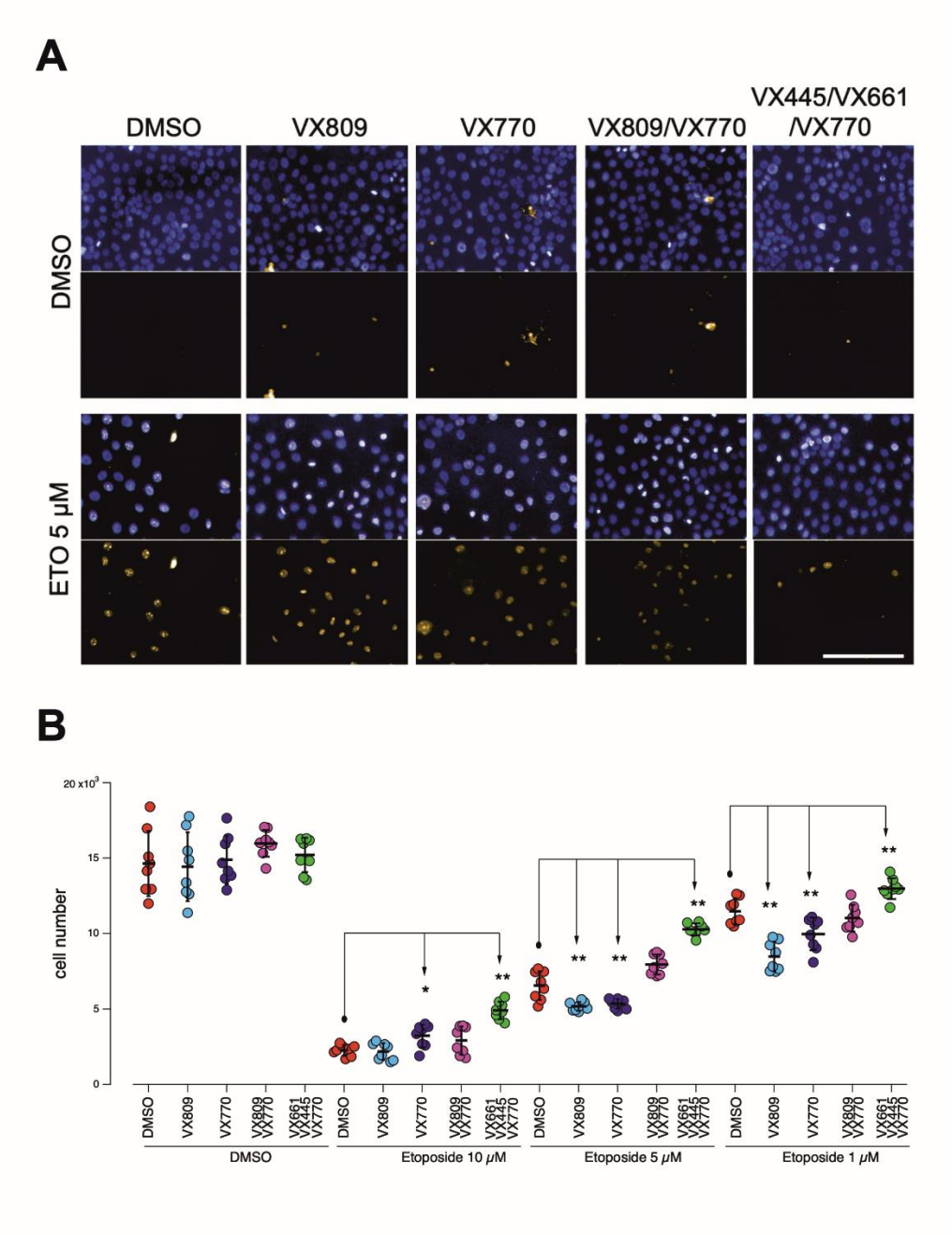


*Figure 51: CFBE41o- cells were treated for 24 hours with vehicle alone or with the indicated concentrations of etoposide to induce apoptosis, in the absence or presence of different CFTR modulators. Samples were then analyzed by automated high-content imaging and analysis. (A) Representative images. Scale bar: 200 µm. (B) Quantification of the number of viable CFBE41o- cells per condition. Data represent mean ± SEM, n = 5. *P < 0.05, **P < 0.01 compared with respective control (DMSO-treated) cells, 1-way ANOVA test with Dunnett's post hoc test. . Figure adapted from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta* [149].

The results of this experiment show that, when cells undergo the pro-apoptotic stimuli, only the treatment with the triple combination of drugs has a protective effect, measured by the higher number of viable cells. On the contrary, cells treated with other CF drugs (either alone or in combination) showed a decrease in the number of viable cells after the pro-apoptotic stimulus. Taken together these results

suggest that the downregulation of the six major ceramides might have a protective role in cell response to pro-apoptotic stimuli.

Beside the ceramides, a downregulation of the three major Lysophosphatidylcholines (LysoPCs) was also detected by the untargeted experiment.

Despite their possible role as biomarkers for CF in some biofluids [219], their exact role in CF is not completely clear yet. LysoPCs are known not only to be secreted to attract macrophages [220] but also to be the precursors of molecules (phosphocholine and glycerophosphocholine) involved in the regulation of the membrane fluidity in mammalian cells [221].

The targeted analysis of LysoPCs conducted (Paragraph 3.4.4) for this work shows that the decrease of this class of lipids is caused only by the treatment with the triple combination of drugs (Figure 52, Panel A).

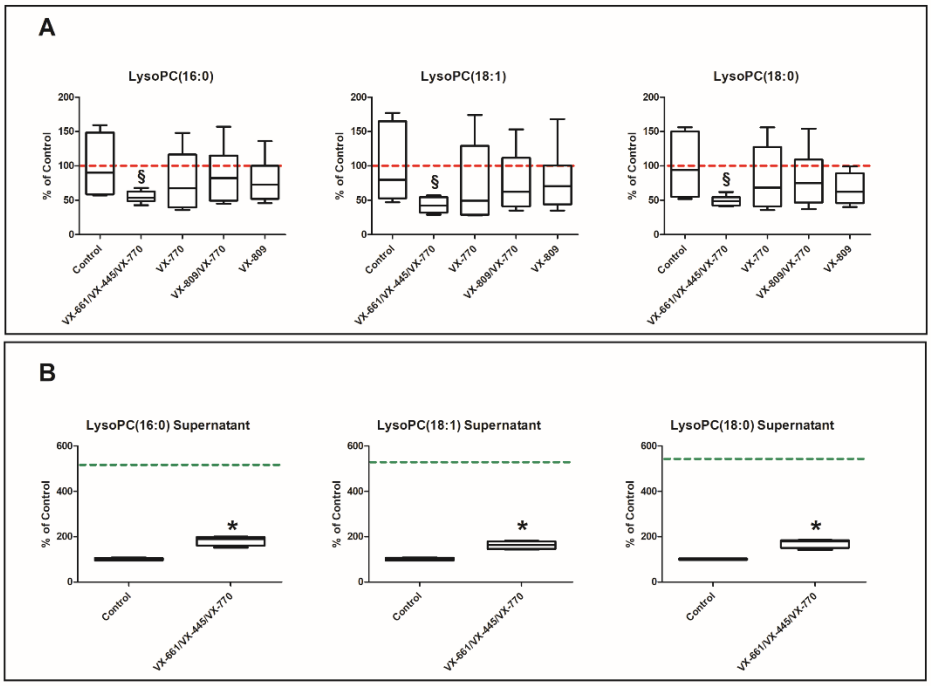


*Figure 52: Targeted quantification of three LysoPCs in the CFBE41o- cells. (A) LysoPC levels in lysates following treatment with the triple combination (VX-661/ VX-445/VX-770) after 24 hours of incubation. The average level of the control group is represented by the dotted red line. §p < 0.05 compared with control, 1-way ANOVA test with Dunnett'st's post hoc test. (B) LysoPC levels in the supernatants of the same CFBE41o- cells at 24 hours, following treatment with the triple combination. The average level in the medium at the beginning of the incubation is indicated by the dotted green line. Box-and-whiskers plot represent the median value (inner bar), the first and third quartile (box), and the minimum and maximum value (outer bars). Data represent mean ± SEM, n = 5. *P < 0.05 compared with control, 2-tailed t test. Figure taken from Liessi N., Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta* [149].

These lipids were also measured in the supernatant of the cells, where their levels are higher if compared to the controls (Figure 52, Panel B). This suggested that the cells, when treated with Kaftrio, reduce the consumption of LysoPCs to produce PCs, which are downregulated together with the LysoPCs in the Kaftrio dataset (Table 11). The biological meaning of these results deserves further investigation.

4.5 LOPIT-DC experiments

4.5.1 Evaluation of the separation of the organelles through WB

In the frame of this work, we also conducted a set of LOPIT-DC (**L**ocalization of **O**rganelle **P**roteins by **I**sotopic **T**agging – **D**ifferential **C**entrifugation, [87]) experiments at the University of Cambridge, in collaboration with Prof. Kathryn Lilley. The biological role and function of a protein is not only related to its level of expression and biochemical activity, but it is also related to ***where*** this activity is carried out within the cell. LOPIT-DC enables the proteome-scale exploration of the spatial distribution and localization of thousands of proteins within the cell, thanks to its ability to separate the different cell organelles by density ultracentrifugation and to label them with specific isotopically coded tags.

Since CF is (also) a trafficking disease, we explored the idea that the LOPIT-DC protocol could be successfully applied to CF research and that this technology could highlight proteins changing their spatial localization in the cell, particularly after CFTR rescue induced by treatment with VX-809. The hypothesis is, again, that omics data helps understanding the alterations in the cell chemical space composition associated to successful CFTR rescue, thus generating knowledge and, hopefully, new pharmacological targets.

During the optimization of LOPIT-DC workflow, Western blotting can help in getting information about the distribution of key organelle markers through the collected fractions (for information about the antibodies used, please refer to Paragraph 3.5.2). After the optimization steps, LOPIT-DC workflow was applied to three biological replicates of CFBE41o- cells after the treatment with VX-809 (1 µM for 24 hours) and their controls. To confirm the efficiency of the organelles separation, for each sample (6 in total, 3 controls and 3 treated) Western-blotting were performed. The corresponding results of the WB are showed in Figure 53.

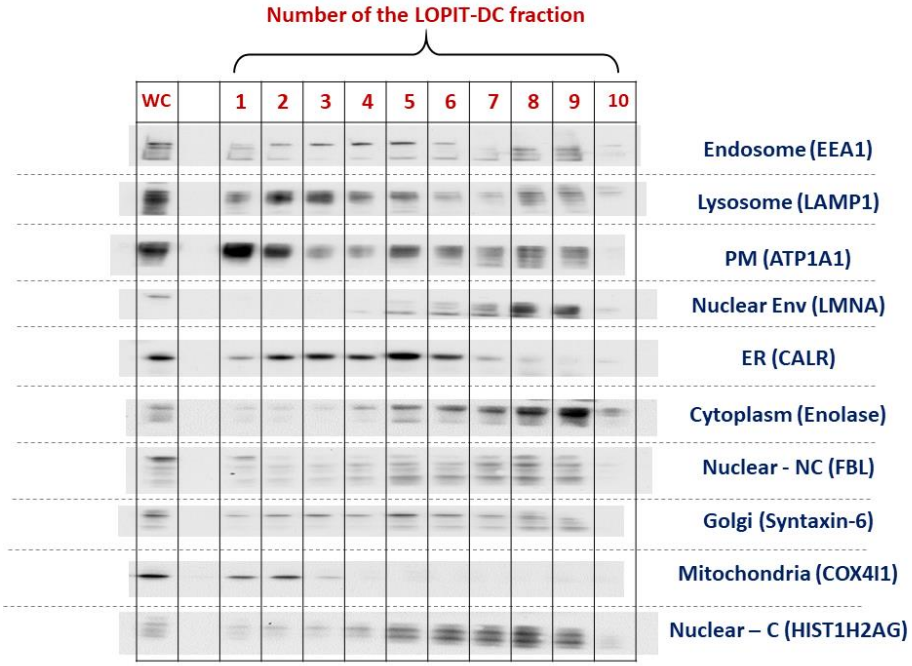


Figure 53: Wester-Blot of Control replicate 1 of CFBE41o- cells after LOPIT-DC centrifugation steps.

This figure shows the profile of each organelle marker in all the fractions collected in the LOPIT-DC experiment. In the LOPIT-DC, several centrifugation steps are performed and a total of ten fractions are normally collected. LOPIT-DC relies on the evidence that proteins from the same subcellular compartment show the same abundance profile across the fractions collected by differential centrifugation. Western blotting is used to evaluate the efficiency in organelle separation, using organelle-specific marker proteins.

To achieve a proper separation of the organelles, each organelle-specific marker protein must show a unique profile distribution along the ten fractions. For some organelles, the separation is clearly detectable. For example, from Figure 53 it is clear that mitochondria proteins are present only in the first two fractions.

Nevertheless, even though for some organelle markers the separation is not as evident as for the mitochondria, their profile distribution is still unique and adequate to achieve the organelle separations. Indeed, ER marker profile is highly enriched in fraction 5 and even though the signal is present in many of the fractions its profile is still unique.

From these results, it is clear that the organelles separation was achieved by applying the LOPIT-DC workflow. As positive control, the whole cells fraction (WC) is also blotted: as expected, a signal for each organelle is detected in the whole cells fraction since in this fraction all the organelles are present all together.

4.5.2 PCA score plots

With LC-MS/MS experiments, more than 5000 proteins were quantified in each sample. The exact numbers of the quantified proteins, after normalization and missing values removal, are summarized in Table 12.

Sample	Total quantified proteins
Control_Rep1	5784
Control_Rep2	6120
Control_Rep3	5941
Treated_Rep1	5975
Treated_Rep2	6059
Treated_Rep3	5940

Table 12: Table summarizing the quantified proteins in each sample after the LC-MS/MS analysis.

After applying the analytical workflow for the spatial proteomics data [181], six PCA score plot maps were built to evaluate the efficiency of the organelle separations in all the samples.

The three replicates were combined together obtaining new maps (one for the controls and one for the treated), showed in Figure 54.

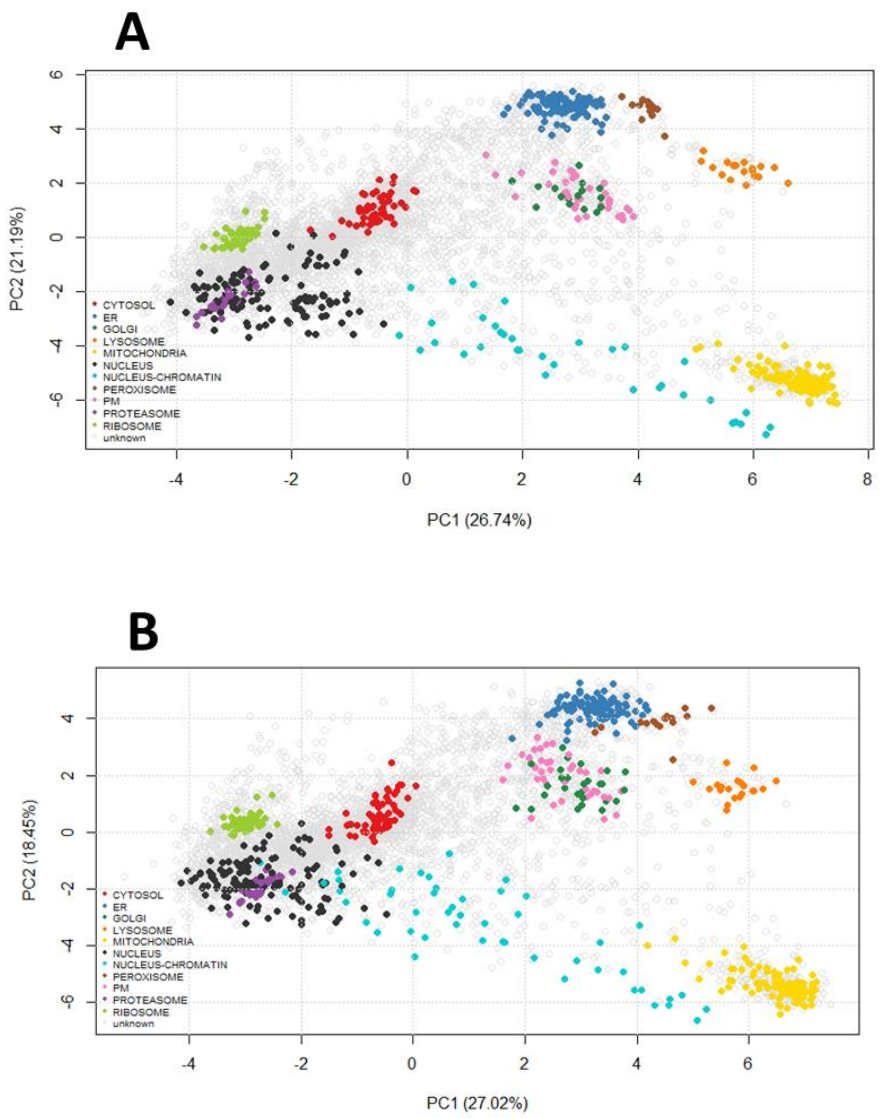


Figure 54: PCA score plots of the combined replicates of each condition. Panel A shows the PCA score plot of the control sample and Panel B shows the plot for the combined treated samples.

Each of these maps represents the final step of the LOPIT-DC analysis. Each dot represents a protein quantified in the sample and its position in the map is linked to its profile distribution along the fractions collected. Proteins with similar profiles are close in the PCA score plot, while proteins with very different profile distribution are distant in the map.

The colored dots are the proteins used as markers for specific organelle (each color corresponds to an organelle), while the grey dots are proteins with unknown localization. From these maps, a clear clusterization of the organelles can be detected, thus indicating that an effective organelle separation was achieved for CFBE41o- cells. While some organelles do not appear perfectly separated over PC1 and PC2, PCA plot over other principal components, like PC3, indicates the separation is visible. As an example, Figure 55 shows the maps of one sample (Control, replicate 3) over other PCs.

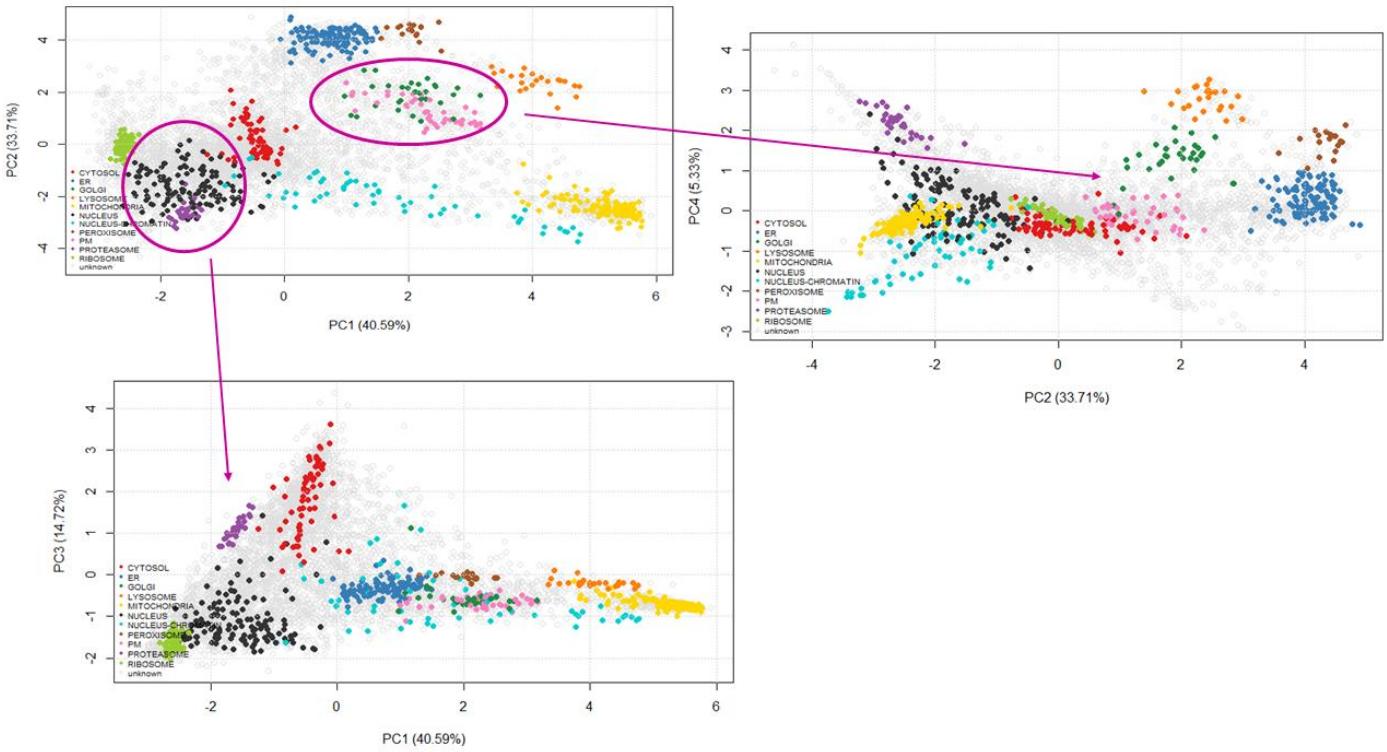


Figure 55: PCA maps of Control, replicate 3 over different PCs. Percentage variance explained by each component is shown in parentheses.

From Figure 55, the separation of nucleus (black dots) and the ribosome (purple dots) is visible over PC1 and PC3, while the separation of Golgi apparatus (dark green dots) and ER (pink dots) is detected over PC2 and PC4.

4.5.3 Data statistics to detect *movers*

In order to introduce data statistics of the three replicates per group in the LOPIT-DC spatial analysis, the SVM classification was performed using the ***4368*** proteins in common between the controls and the treated samples following a dedicated workflow for spatial proteomics data [181].

The numbers of the proteins assigned to each organelle are shown in Table 13.

Organelle	Controls	Treated
Cytosol	581	561
ER	264	269
Golgi	48	43
Lysosome	34	34
Mitochondria	197	197
Nucleus	809	836
Nucleus-chromatin	127	134
Peroxisome	25	25
PM	158	154
Proteasome	49	48
Ribosome	168	162
Unknown	1908	1905

Table 13: Summary of proteins having an organelle-specific localization after the SVM classification.

Another statistical analysis was performed [89] to identify the movers, those proteins that change their localization after the treatment.

4.5.4 The movers

To detect proteins that changed their localization after the treatment with VX-809 ("***movers***"), TAGM (t-augmented Gaussian mixture) statistical analysis was performed, following a well-established workflow [89]. From the TAGM analysis, **231** movers were identified. A score, directly proportional to the difference between the profiles of the controls and the treated samples, was calculated for each of these proteins.

Interestingly, CFTR was not detected among the movers. Indeed, CFTR was detected in ER in both the control and the treated samples. This result is not surprising since F508del-CFTR resides mainly in the ER (as opposed to the wt-CFTR, which is more present at the PM) [34]. Moreover, these cells were created and selected to stably overexpress only F508del-CFTR, whose primary localization is the ER. LOPIT-DC, as several profiling spatial proteomics methods, detects only the main (or primary) localization of the proteins - the subcellular niche where the proteins are resident for most of their life cycle [222]. LOPIT-DC workflow is not able to detect the relatively small amount of CFTR (less than 30%) known to be rescued by VX-809, and yet able to elicit a pharmacologically relevant effect on the CF condition [223], [148].

Among the proteins classified as movers, 6 CFTR interactors were detected. As shown in Table 14, the Bayes Factor associated to these proteins is quite low, suggesting that their subcellular localization is not dramatically changed after the treatment.

Accession	Protein names	Gene	Bayes Factor
Q9Y385	Ubiquitin-conjugating enzyme E2 J1 (EC 2.3.2.23) (E2 ubiquitin-conjugating enzyme J1) (Non-canonical ubiquitin-conjugating enzyme 1) (NCUBE-1) (Yeast ubiquitin-conjugating enzyme UBC6 homolog E) (HsUBC6e)	UBE2J1	95.83
Q9UBV2	Protein sel-1 homolog 1 (Suppressor of lin-12-like protein 1) (Sel-1L)	SEL1L	80.37
B4DR61	Protein transport protein Sec61 subunit alpha isoform 1 (cDNA FLJ59739, highly similar to Protein transport protein Sec61 subunit alpha isoform 1)	SEC61A1	60.71
O43520	Phospholipid-transporting ATPase IC (EC 7.6.2.1) (ATPase class I type 8B member 1) (Familial intrahepatic cholestasis type 1) (P4-ATPase flippase complex alpha subunit ATP8B1)	ATP8B1	44.23
J3KPS0	DnaJ (Hsp40) homolog, subfamily B, member 12, isoform CRA_c (DnaJ homolog subfamily B member 12)	DNAJB12	42.15
P17612	cAMP-dependent protein kinase catalytic subunit alpha (PKA C-alpha) (EC 2.7.11.11)	PRKACA	41.41

Table 14: List of CFTR interactors among the proteins classified as movers after the treatment with the VX-809.

Moreover, proteins involved in protein transport were detected among the movers and they are listed in Table 15.

Accession	Protein names	Gene	Bayes Factor
P35914	Hydroxymethylglutaryl-CoA lyase, mitochondrial (HL) (HMG-CoA lyase) (EC 4.1.3.4) (3-hydroxy-3-methylglutarate-CoA lyase)	HMGCL	286.94
A8MXV4	Nucleoside diphosphate-linked moiety X motif 19 (Nudix motif 19) (EC 3.6.1.-)	NUDT19	145.55
Q86WA8	Lon protease homolog 2, peroxisomal (EC 3.4.21.53) (Lon protease-like protein 2) (Lon protease 2) (Peroxisomal Lon protease) (pLon)	LONP2	129.46
P12235	ADP/ATP translocase 1 (ADP,ATP carrier protein 1) (ADP,ATP carrier protein, heart/skeletal muscle isoform T1) (Adenine nucleotide translocator 1) (ANT 1) (Solute carrier family 25 member 4)	SLC25A4	90.87
Q9NS69	Mitochondrial import receptor subunit TOM22 homolog (hTom22) (1C9-2) (Translocase of outer membrane 22 kDa subunit homolog)	TOMM22	86.07
P12236	ADP/ATP translocase 3 (ADP,ATP carrier protein 3) (ADP,ATP carrier protein, isoform T2) (ANT 2) (Adenine nucleotide translocator 3) (ANT 3) (Solute carrier family 25 member 6) [Cleaved into: ADP/ATP translocase 3, N-terminally processed]	SLC25A6	73.44
C9JRZ6	MICOS complex subunit	CHCHD3	57.18
Q16595	Frataxin, mitochondrial (EC 1.16.3.1) (Friedreich ataxia protein) (Fxn) [Cleaved into: Frataxin intermediate form (i-FXN); Frataxin(56-210) (m56-FXN); Frataxin(78-210) (d-FXN) (m78-FXN); Frataxin mature form (Frataxin(81-210)) (m81-FXN)]	FXN	51.19
B7Z2R7	Acyl-CoA-binding domain-containing protein 5	ACBD5	49.98
Q3ZCQ8	Mitochondrial import inner membrane translocase subunit TIM50	TIMM50	48.20
O43808	Peroxisomal membrane protein PMP34 (34 kDa peroxisomal membrane protein) (Solute carrier family 25 member 17)	SLC25A17	44.72
A0A0C4DGA2	Enoyl-CoA delta isomerase 2, mitochondrial (Testicular secretory protein Li 33)	ECI2	44.68
Q92968	Peroxisomal membrane protein PEX13 (Peroxin-13)	PEX13	44.37

Table 15: Proteins involved in protein transport among the movers detected by LOPIT-DC experiments.

The full paper reporting the whole set of results for the application of LOPIT-DC to CFBE41o- cells is almost ready. The last set of validation experiments is ongoing and the manuscript will be submitted soon.

5 Discussion

My research work deals with the use of advanced MS-based omics tools to further investigate the pathophysiology of cystic fibrosis and to highlight changes in the chemical space composition of CFBE41o- cells resulting from successful CFTR rescue maneuvers. The final aim of this effort is to generate data useful to uncover new biological mechanisms contributing to CF disease and, hopefully, to identify new potential targets for the pharmacological treatment of CF. In the present work, several MS-based omics techniques were applied to gain an insight into the molecular mechanisms associated with the rescue of the mutant CFTR at the plasma membrane.

First, and in view of the subsequent experimental activities, a new tool for label-free proteomics workflow was developed and optimized, making this powerful MS-based protocol perfectly suitable for CF research. We deeply analyzed the proteome of both immortalized and primary bronchial epithelium cells following several fractionation strategies in order to identify proteins specifically expressed in bronchial epithelium cells. We added to the Pan human ion library – the most complete ion library available so far for human samples - assays to quantify CFTR and other proteins known to be involved in CF pathology.

With this new resource in hand, a deep proteomic profiling of BE primary cells was performed, by comparing the proteomes of four CF patients with four non-CF subjects, used as controls. We aimed at demonstrating the suitability of our newly-developed ion library for CF research. Moreover, we wanted to study and monitor changes in protein expression induced by CF pathology in well-differentiated bronchial epithelial cells. Many proteins (and pathways) were found to be significantly dysregulated by the disease. Several of these proteins are known to be direct CFTR interactors or proteins involved in CFTR trafficking and degradation. This study led to the publication of my first paper as first author in 2018 in the Journal of Cystic Fibrosis [151].

A similar approach to identify proteins with a differential expression in CF was performed both by Bini's group in 2018 [112] and by Yate's group in 2014 [224]: the proteome of CFBE41o- cells (expressing F508del-CFTR) was compared to that of 16HBE14o- cells (expressing wt-CFTR). Despite the use of different quantitative approaches and different nature of the samples, we found some proteins dysregulated in CF in common with these studies, such as the tubulin alpha-1B chain (TUBA1B) and the protein S100-A6. It is important to keep in mind that data obtained with immortalized cell lines may not be representative of the *in-vivo* phenotype. For this reason, the dysregulation of the proteins in common between these studies can suggest their potential crucial role in CF biology.

Despite the limited number of subjects, we believe that our work represents an improvement in the investigation of primary cells derived from CF patients. The lack of CFTR among the quantified protein was somehow expected, given its low natural abundance in human bronchial epithelium cells. On the other hand, our newly-developed IL enables us to quantify CFTR in all the experiments conducted using CFBE41o- cells: this cell model was created for the purpose of over-expressing the mutant channel; therefore it comes as no surprise that we detected CFTR in these cells.

The levels of F508del-CFTR were found to be significantly increased in association with the gene silencing of the four "primary targets" – proteins that, when silenced are known to trigger mutant CFTR rescue at the plasma membrane [148]. The significant increase in CFTR levels following genetic maneuvers is an expected result: the proteins whose expressions were inhibited are known (or supposed to) interfere with F508del-CFTR degradation and aimed at increasing the quantity of the mutant CFTR available in the cell.

We used this approach to highlight proteins whose levels of expressions are altered in association with CFTR rescue. Our aim with this experiment was to gain a crucial understanding of how the cell reacts to F508del-CFTR rescue. Our strategy was to use siRNA to selectively silence four primary targets (known to trigger CFTR rescue) in order to detect the proteins and the pathways significantly altered by this genetic CFTR rescue. This approach allowed us to identify a list of putative new targets to be further investigated. We then used data mining strategies and tools to select molecules able to modulate these targets and we tested them on CFBE41o- cells to assess their effect on CFTR rescue. Some of these compounds indeed succeeded in inducing CFTR rescue. On the other hand, the only compound that we tested on bronchial epithelium cells so far failed to show the same effect on primary cells. Due to a set of unexpected delays, mainly related to the 2020 COVID-19 pandemic, further tests on the remaining compounds are still pending. In any case, as expected from the beginning of the project, further efforts are still needed to translate CFTR rescue from CFBE41o- to primary cells. Indeed, early validation of primary hits in a native cell model is a crucial step in CF drug development, due to the demonstrated influence of cell background on the pharmacological rescue of mutant CFTR, in particular for compounds acting as correctors of the maturation defect [225], [226].

Beside the genetic rescue of CFTR obtained by siRNA, we also investigated the alterations associated with a pharmacological rescue. Although several recent studies were published to highlight the importance of membrane lipids in the physiology and trafficking of CFTR, an untargeted lipidomics profiling of bronchial epithelial cells was never performed. For this reason, we investigated how CF drugs alter the CFBE41o- lipidome. We hypothesized that the pharmacological rescue of mutant CFTR could be associated with changes in lipidome profile of CFBE41o- cells.

In this work, the first untargeted lipidome profiling of CFBE41o- was conducted to study the effect of different pharmacological treatments on lipids. This study led a publication of a paper in the JCI Insight journal [149]. The results of this experiment demonstrated that each of the tested drugs triggers a characteristic change in the cell lipidome. We found that VX-770 is associated with a particular lipid signature, consisting in the decrease of 7 phosphatidylcholines and a ceramide. Moreover, we discovered that the treatment with Kaftrio induces a significant decrease of 6 ceramides. This effect is also associated with an increased resistance of treated CFBE41o- to pro-apoptotic stimuli. These findings represent a first step for future research to define the role of lipids, and sphingolipids in particular, in mutant CFTR rescue. Indeed, in the last few years, the crucial role of membrane lipids in CF was investigated. In particular, cholesterol influences the fluidity of the plasma membrane, as it forms lipid clusters and rafts together with sphingomyelin [227]. This topic is important in CF research as in murine tracheal epithelial cells CFTR was found to colocalize with these lipid rafts [228]; the presence of CFTR in lipid rafts is essential during *P. aeruginosa* infection [229]. Given that the surrounding lipid environment can affect CFTR activity, studying the imbalances of lipids in people with CF can shed light on their role in CF biology. Moreover, the modulators of CFTR activity can act at their full potential if CFTR is in a proper lipid environment. For this reason, we think that our untargeted profiling of changes in lipids associated with different pharmacological treatments can be of help to future research in this field.

Finally, in the last part of my work, the LOPIT-DC workflow was applied to the CFBE41o- cells. LOPIT-DC is the current state-of-the-art tool to simultaneous investigate the subcellular localization of the proteins from complex biological mixtures in a single experiment.

With the help of Prof. Lilley's group at the University of Cambridge, LOPIT-DC workflow was optimized to obtain the first subcellular protein localization map of CFBE41o- cells. We then investigated if a pharmacological F508del-CFTR rescue (obtained with the corrector VX-809) alters the subcellular localization of proteins in CFBE41o- cells. We found that the treatment with the corrector VX-809 is associated with changes in subcellular localization of more than 200 proteins, defined as "movers". The detected movers could be involved in the functional rescue of the mutant CFTR or related to CFTR activity and trafficking. We believe that the identification of these "movers" can help in further understanding of the CF pathophysiological mechanisms and perhaps help the quest for new pharmacological targets.

Many different versions of LOPIT protocol have been published [78]–[83], [87] but this method was often applied to a "static" situation: the subcellular protein localization was investigated using only a single condition. For the first time, with our experiment, LOPIT-DC was applied to CF research. Moreover, we studied a "dynamic situation", comparing two different conditions in order to detect those proteins that changed their subcellular localization in association with the pharmacological rescue of the mutant CFTR.

Since CF is (also) a trafficking disease, we explored the idea that the LOPIT-DC protocol could be successfully applied to CF research and that this technology could highlight proteins changing their spatial localization in the cell, particularly after CFTR rescue induced by treatment with VX-809. The hypothesis is, again, that omics data helps understanding the alterations in the cell chemical space composition associated to successful CFTR rescue, thus generating knowledge and, hopefully, new pharmacological targets. A paper describing the results obtained from the LOPIT-DC experiment is under preparation.

6 Scientific publications

6.1 First year

- A new SWATH ion library for mouse adult hippocampal neural stem cells. (Data Brief. 2018, 18: 1–8) Authors: **Braccia, C**., Espinal, M.P., Pini, M., De Pietri Tonelli, D., Armirotti, A. doi: 10.1016/j.dib.2018.02.062

6.2 Second year

- MiR-135a-5p Is Critical for Exercise-Induced Adult Neurogenesis. (Stem Cell Reports. 2019 (6):1298-1312) Authors: Pons-Espinal, M., Gasperini, C., Marzi, M.J., **Braccia, C**., Armirotti, A., Pötzsch, A., Walker, T.L., Fabel, K.d,e, Nicassio, F., Kempermann, G., De Pietri Tonelli, D. doi: 10.1016/j.stemcr.2019.04.020
- Comparative Proteomic Analysis of Proteins Involved in Bioenergetics Pathways Associated with Human Sperm Motility. (Int J Mol Sci. 2019, 20(12):3000) Authors: Moscatelli, N., Lunetti, P., **Braccia, C**., Armirotti, A., Pisanello, F., De Vittorio, M., Zara, V., Ferramosca, A. doi: 10.3390/ijms20123000
- SWATH label-free proteomics for cystic fibrosis research. (J Cyst Fibros. 2019, 18(4):501-506) Authors: **Braccia, C**., Tomati, V., Caci, E., Pedemonte, N., Armirotti, A. doi: 10.1016/j.jcf.2018.10.004
- Bioactive Thymosin Alpha-1 Does Not Influence F508del-CFTR Maturation and Activity. (Sci Rep. 2019, 9(1):10310) Authors: Armirotti A., Tomati, V., Matthes, E., Veit, G., Cholon, D.M., Phuan, P.W., **Braccia, C**. Guidone, D. Gentzsch, M. Lukacs, G.L. Verkman, A.S. Galietta, L. J. V. Hanrahan, J.W. & Pedemonte, N. doi: 10.1038/s41598-019-46639-1

6.3 Third year

- Proteomics and metabolomics for cystic fibrosis research. (Int J Mol Sci. 2020, 21(15):5439) Authors: Liessi, N., Pedemonte, N., Armirotti, A., **Braccia, C**. doi: 10.3390/ijms21155439
- Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta. (JCI Insight. 2020, 5(16):e138722) Authors: Liessi N., Pesce E., **Braccia C**., Bertozzi S. M., Giraudo A., Bandiera T., Pedemonte N., Armirotti A. doi: 10.1172/jci.insight.138722
- Proteomics analysis of FUS mutant human motoneurons reveals altered regulation of cytoskeleton and other ALS-linked proteins via 3'UTR binding. (Sci Rep. 2020, 10(1):11827) Authors: Garone, M.G., Alfano, V., Salvatori, B., **Braccia, C.**, Peruzzi, G., Colantoni, A., Bozzoni, I., Armirotti, A., Rosa, A. doi: 10.1038/s41598-020-68794-6
- GADD34 is a modulator of autophagy during starvation (Sci Adv. 2020, 6(39):eabb0205) Authors: Gambardella G., Staiano L., Moretti M. N., De Cegli R., Fagnocchi L., Di Tullio G., Polletti S., **Braccia C.**, Armirotti A., Zippo A., Ballabio A., De Matteis M. A., di Bernardo D. doi: 10.1126/sciadv.abb0205
- Book chapter: Proteomics Data Analysis, published by Springer Nature. "Quantification of changes in protein expression using swath proteomics. Authors: **Braccia C**., Liessi N., Armirotti A. ***in press***

7 Acknowledgments

These works were supported by Italian Cystic Fibrosis Foundation grants (FFC 1/2018 and FFC 1/2019 to Andrea Armirotti and FFC 9/2019 to Nicoletta Pedemonte) and contributions from “Delegazione FFC di Genova con Gruppo di sostegno FFC di Savona Spotorno,” “Delegazione FFC di Valle Scrivia Alessandria,” “Delegazione FFC di Montescaglioso,” and “Delegazione FFC di Ascoli Piceno.”

Moreover, most of these experiments were conducted in collaboration with the Dr. Pedemonte’s lab at the Istituto Giannina Gaslini, Genova, Italy. The work in Dr. Pedemonte’s lab is also supported by the Italian Ministry of Health, through Cinque per mille and Ricerca Corrente (Linea 1).

All the LOPIT-DC experiments were conducted thanks to the collaboration with Professor Lilley’s lab at the University of Cambridge.

I would like to acknowledge all the people who worked with me during these three years.

First of all, I would like to thank my supervisor Dr. Armirotti for all the teachings and the opportunities he gave to me: he was an excellent mentor for me.

I would like to acknowledge my second supervisor, Dr. Bandiera for his support and availability. Moreover, I would like to thank all the people I met in the Analytical Chemistry Lab in IIT: Giuliana, Sine, Nara and Sion. I would like to mention also Dr. Pedemonte and Dr. Tomati at Gaslini Hospital Lab for teaching me the cell culture.

I would also thank my family – my dad, my uncles, my grandparents, my sister and my boyfriend - for their love and support during these three years.

In the end, I have to thank the woman who is my guidepost for everything: my mom.

8 Bibliography

[1] Cystic Fibrosis Foundation, “Patient Registry Annual Data Report,” 2018.

[2] D. H. ANDERSEN, “CYSTIC FIBROSIS OF THE PANCREAS AND ITS RELATION TO CELIAC DISEASE: A CLINICAL AND PATHOLOGIC STUDY,” *Am. J. Dis. Child.*, vol. 56, no. 2, pp. 344–399, Aug. 1938, doi: 10.1001/archpedi.1938.01980140114013.

[3] P. A. DI SANT’AGNESE, R. C. DARLING, G. A. PERERA, and E. SHEA, “ABNORMAL ELECTROLYTE COMPOSITION OF SWEAT IN CYSTIC FIBROSIS OF THE PANCREAS,” *Pediatrics*, vol. 12, no. 5, pp. 549 LP – 563, Nov. 1953, [Online]. Available: http://pediatrics.aappublications.org/content/12/5/549.abstract.

[4] J. Bijman and P. M. Quinton, “Influence of abnormal Cl- impermeability on sweating in cystic fibrosis,” *Am. J. Physiol. - Cell Physiol.*, vol. 16, no. 1, 1984, doi: 10.1152/ajpcell.1984.247.1.c3.

[5] J. Bijman and P. M. Quinton, “Apparent absence of cystic fibrosis sweat factor on ion–selective and transport properties of the perfused human sweat duct,” *Pediatr. Res.*, vol. 18, no. 12, pp. 1292–1296, 1984, doi: 10.1203/00006450-198412000-00014.

[6] M. R. Knowles, M. J. Stutts, A. Spock, N. Fischer, J. T. Gatzy, and R. C. Boucher, “Abnormal ion permeation through cystic fibrosis respiratory epithelium,” *Science (80-.).*, vol. 221, no. 4615, pp. 1067–1070, 1983, doi: 10.1126/science.6308769.

[7] M. P. Rogan, D. A. Stoltz, and D. B. Hornick, “Cystic Fibrosis Transmembrane Conductance Regulator Intracellular Processing, Trafficking, and Opportunities for Mutation-Specific Treatment,” *Chest*, vol. 139, no. 6, pp. 1480–1490, Jun. 2011, doi: 10.1378/chest.10-2077.

[8] J. R. Riordan *et al.*, “Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA,” *Science (80-.).*, vol. 245, no. 4922, pp. 1066–1073, 1989, doi: 10.1126/science.2475911.

[9] T. C. Hwang, J. T. Yeh, J. Zhang, Y. C. Yu, H. I. Yeh, and S. Destefano, “Structural mechanisms of CFTR function and dysfunction,” *J. Gen. Physiol.*, vol. 150, no. 4, pp. 539–570, 2018, doi: 10.1085/jgp.201711946.

[10] T. F. Scanlin and M. C. Glick, “Terminal glycosylation in cystic fibrosis,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1455, no. 2–3, pp. 241–253, 1999, doi: 10.1016/S0925-4439(99)00059-9.

[11] P. Vergani, S. W. Lockless, A. C. Nairn, and D. C. Gadsby, “CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains,” *Nature*, vol. 433, no. 7028, pp. 876–880, 2005, doi: 10.1038/nature03313.

[12] D. C. Gadsby and A. C. Nairn, “Control of CFTR channel gating by phosphorylation and nucleotide hydrolysis,” *Physiol. Rev.*, vol. 79, no. 1 SUPPL. 1, 1999, doi: 10.1152/physrev.1999.79.1.S77.

[13] M. Cohen-Cymberknoh, D. Shoseyov, and E. Kerem, “Managing cystic fibrosis: Strategies that increase life expectancy and improve quality of life,” *Am. J. Respir. Crit. Care Med.*, vol. 183, no. 11, pp. 1463–1471, 2011, doi: 10.1164/rccm.201009-1478CI.

[14] S. Blouquit *et al.*, “Ion and fluid transport properties of small airways in cystic fibrosis,” *Am. J. Respir. Crit. Care Med.*, vol. 174, no. 3, pp. 299–305, 2006, doi: 10.1164/rccm.200506-987OC.

[15] P. Duchesneau, T. K. Waddell, and G. Karoubi, “Cell-Based Therapeutic Approaches for Cystic Fibrosis,” no. 2018.

[16] “Clinical and Functional Translation of CFTR (CFTR2) www.cftr2.org.” .

[17] N. Derichs, “Targeting a genetic defect: Cystic fibrosis transmembrane conductance regulator modulators in cystic fibrosis,” *Eur. Respir. Rev.*, vol. 22, no. 127, pp. 58–65, 2013, doi: 10.1183/09059180.00008412.

[18] K. De Boeck and M. D. Amaral, “Progress in therapies for cystic fibrosis,” *Lancet Respir. Med.*, vol. 4, no. 8, pp. 662–674, 2016, doi: 10.1016/S2213-2600(16)00023-0.

[19] M. Genetics and S. Diego, “Worldwide survey of the delta F508 mutation--report from the cystic fibrosis genetic analysis consortium.,” *Am. J. Hum. Genet.*, vol. 47, no. 2, pp. 354–9, 1990, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2378364%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1683705.

[20] G. L. Lukacs *et al.*, “The AF508 Mutation Decreases the Stability of Cystic Fibrosis Transmembrane Conductance Regulator in the Plasma Membrane,” *J. Biol. Chem.*, vol. 268, no. 29, pp. 21592–21598, 1993.

[21] V. Kaminskyy and B. Zhivotovsky, “Proteases in autophagy,” *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1824, no. 1, pp. 44–50, 2012, doi: 10.1016/j.bbapap.2011.05.013.

[22] J. J. Caramelo and A. J. Parodi, “How sugars convey information on protein conformation in the endoplasmic reticulum,” *Semin. Cell Dev. Biol.*, vol. 18, no. 6, pp. 732–742, 2007, doi: 10.1016/j.semcdb.2007.09.006.

[23] C. M. Farinha, P. Matos, and M. D. Amaral, “Control of cystic fibrosis transmembrane conductance regulator membrane trafficking: Not just from the endoplasmic reticulum to the Golgi,” *FEBS J.*, vol. 280, no. 18, pp. 4396–4406, 2013, doi: 10.1111/febs.12392.

[24] J. W. Hanrahan, H. M. Sampson, and D. Y. Thomas, “Novel pharmacological strategies to treat cystic fibrosis,” *Trends Pharmacol. Sci.*, vol. 34, no. 2, pp. 119–125, 2013, doi: 10.1016/j.tips.2012.11.006.

[25] S. Vashist and D. T. W. Ng, “Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control,” *J. Cell Biol.*, vol. 165, no. 1, pp. 41–52, 2004, doi: 10.1083/jcb.200309132.

[26] C. M. Farinha, P. Nogueira, F. Mendes, D. Penque, and M. D. Amaral, “The human Dnaj homologue (Hdj)-1/heat-shock protein (Hsp) 40 co-chaperone is required for the in vivo stabilization of the cystic fibrosis transmembrane conductance regulator by Hsp70,” *Biochem. J.*, vol. 366, no. 3, pp. 797–806, 2002, doi: 10.1042/BJ20011717.

[27] P. Määttänen, K. Gehring, J. J. M. Bergeron, and D. Y. Thomas, “Protein quality control in the ER: The recognition of misfolded proteins,” *Semin. Cell Dev. Biol.*, vol. 21, no. 5, pp. 500–511, 2010, doi: 10.1016/j.semcdb.2010.03.006.

[28] M. Roxo-Rosa *et al.*, “Revertant mutants G550E and 4RK rescue cystic fibrosis mutants in the first nucleotide-binding domain of CFTR by different mechanisms,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 47, pp. 17891–17896, 2006, doi: 10.1073/pnas.0608312103.

[29] X. Wang *et al.*, “COPII-dependent export of cystic fibrosis transmembrane conductance regulator from the ER uses di-acidic exit code,” *J. Cell Biol.*, vol. 167, no. 1, pp. 65–74, 2004, doi: 10.1083/jcb.200401035.

[30] M. D. Amaral, “CFTR and chaperones: Processing and degradation,” *J. Mol. Neurosci.*, vol. 23, no. 1–2, pp. 41–48, 2004, doi: 10.1385/jmn:23:1-2:041.

[31] S. I. Bannykh, G. I. Bannykh, K. N. Fish, B. D. Moyer, J. R. Riordan, and W. E. Balch, “Traffic pattern of cystic fibrosis transmembrane regulator through the early exocytic pathway,” *Traffic*, vol. 1, no. 11, pp. 852–870, 2000, doi: 10.1034/j.1600-0854.2000.011105.x.

[32] J. S. Yoo, B. D. Moyer, S. Bannykh, H. M. Yoo, J. R. Riordan, and W. E. Balch, “Non-conventional trafficking of the cystic fibrosis transmembrane conductance regulator through the early secretory pathway,” *J. Biol. Chem.*, vol. 277, no. 13, pp. 11401–11409, 2002, doi: 10.1074/jbc.M110263200.

[33] C. M. Farinha and S. Canato, “From the endoplasmic reticulum to the plasma membrane: mechanisms of CFTR folding and trafficking,” *Cell. Mol. Life Sci.*, vol. 74, no. 1, pp. 39–55, 2017, doi: 10.1007/s00018-016-2387-7.

[34] B. N.A., “Intracellular CFTR: Localization and function,” *Physiol. Rev.*, vol. 79, no. 1 SUPPL. 1, pp. S175–S191, 1999, [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed4&NEWS=N&AN=1999045369.

[35] M. Gentzsch *et al.*, “Endocytic trafficking routes of wild type and ΔF508 cystic fibrosis transmembrane conductance regulator,” *Mol. Biol. Cell*, vol. 15, no. 6, pp. 2684–2696, 2004, doi: 10.1091/mbc.E04-03-0176.

[36] D. M. Cholon, W. K. O’Neal, S. H. Randell, J. R. Riordan, and M. Gentzsch, “Modulation of endocytic trafficking and apical stability of CFTR in primary human airway epithelial cultures,” *Am. J. Physiol. - Lung Cell. Mol. Physiol.*, vol. 298, no. 3, 2010, doi: 10.1152/ajplung.00016.2009.

[37] L. S. Prince *et al.*, “Efficient endocytosis of the cystic fibrosis transmembrane conductance regulator requires a tyrosine-based signal,” *J. Biol. Chem.*, vol. 274, no. 6, pp. 3602–3609, 1999, doi: 10.1074/jbc.274.6.3602.

[38] T. Okiyoneda and G. L. Lukacs, “Cell surface dynamics of CFTR: The ins and outs,” *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1773, no. 4, pp. 476–479, 2007, doi: 10.1016/j.bbamcr.2007.01.004.

[39] T. Okiyoneda *et al.*, “Peripheral protein quality control removes unfolded CFTR from the plasma membrane,” *Science (80-.).*, vol. 329, no. 5993, pp. 805–810, 2010, doi: 10.1126/science.1191542.

[40] A. M. Rojas, G. Fuentes, A. Rausell, and A. Valencia, “The Ras protein superfamily: Evolutionary tree and role of conserved amino acids,” *J. Cell Biol.*, vol. 196, no. 2, pp. 189–201, 2012, doi: 10.1083/jcb.201103008.

[41] F. A. Barr, “Rab GTPase function in Golgi trafficking,” *Semin. Cell Dev. Biol.*, vol. 20, no. 7, pp. 780–783, 2009, doi: 10.1016/j.semcdb.2009.03.007.

[42] N. Ameen, M. Silvis, and N. A. Bradbury, “Endocytic trafficking of CFTR in health and disease,” *J. Cyst. Fibros.*, vol. 6, no. 1, pp. 1–14, 2007, doi: 10.1016/j.jcf.2006.09.002.

[43] A. C. Engevik and J. R. Goldenring, “Trafficking ion transporters to the apical membrane of polarized intestinal enterocytes,” *Cold Spring Harb. Perspect. Biol.*, vol. 10, no. 1, pp. 1–16, 2018, doi: 10.1101/cshperspect.a027979.

[44] V. V Ozols, A. Choudhury, and R. E. Pagano, “Endocytic Trafficking Routes of Wild Type and F508del-

CFTR," *Mol. Biol. Cell*, vol. 15, no. June, pp. 2684–2696, 2004, doi: 10.1091/mbc.E04.

[45] J. S. Elborn, "Cystic fibrosis," *Lancet*, vol. 388, no. 10059, pp. 2519–2531, Nov. 2016, doi: 10.1016/S0140-6736(16)00576-6.

[46] T. Okiyoneda *et al.*, "Mechanism-based corrector combination restores ΔF508-CFTR folding and function," *Nat. Chem. Biol.*, vol. 9, no. 7, pp. 444–454, Jul. 2013, doi: 10.1038/nchembio.1253.

[47] K. De Boeck, "Cystic fibrosis in the year 2020: A disease with a new face," *Acta Paediatr. Int. J. Paediatr.*, vol. 109, no. 5, pp. 893–899, 2020, doi: 10.1111/apa.15155.

[48] D. M. Cholon, C. R. Esther, and M. Gentzsch, "Efficacy of lumacaftor-ivacaftor for the treatment of cystic fibrosis patients homozygous for the F508del-CFTR mutation," *Expert Rev. Precis. Med. Drug Dev.*, vol. 1, no. 3, pp. 235–243, May 2016, doi: 10.1080/23808993.2016.1175299.

[49] P. G. Middleton *et al.*, "Elexacaftor-tezacaftor-ivacaftor for cystic fibrosis with a single Phe508del allele," *N. Engl. J. Med.*, vol. 381, no. 19, pp. 1809–1819, 2019, doi: 10.1056/NEJMoa1908639.

[50] B. H. Rosen *et al.*, "Animal and model systems for studying cystic fibrosis," *J. Cyst. Fibros.*, vol. 17, no. 2, pp. S28–S34, 2018, doi: 10.1016/j.jcf.2017.09.001.

[51] G. M. Lavelle, M. M. White, N. Browne, N. G. McElvaney, and E. P. Reeves, "Animal Models of Cystic Fibrosis Pathology: Phenotypic Parallels and Divergences," *Biomed Res. Int.*, vol. 2016, 2016, doi: 10.1155/2016/5258727.

[52] G. Kaur and J. M. Dufour, "Cell lines: Valuable tools or useless artifacts.," *Spermatogenesis*, vol. 2, no. 1, pp. 1–5, 2012, doi: 10.4161/spmg.19885.

[53] E. Bruscia, F. Sangiuolo, P. Sinibaldi, K. K. Goncz, and D. C. Gruenert, "Isolation of CF cell lines corrected at ΔF508-CFTR locus by SFHR-mediated targeting," *Gene Ther.*, vol. 9, no. 11, pp. 683–685, 2002, doi: 10.1038/sj.gt.3301741.

[54] B. Illek, R. Maurisse, L. Wahler, K. Kunzelmann, H. Fischer, and D. C. Gruenert, "Cl transport in complemented CF bronchial epithelial cells correlates with CFTR mRNA expression levels," *Cell. Physiol. Biochem.*, vol. 22, no. 1–4, pp. 57–68, 2008, doi: 10.1159/000149783.

[55] J. R. Dorin *et al.*, "Cystic fibrosis in the mouse by targeted insertional mutagenesis," *Nature*, vol. 359, no. 6392, pp. 211–215, 1992, doi: 10.1038/359211a0.

[56] C. Guilbault, Z. Saeed, G. P. Downey, and D. Radzioch, "Cystic fibrosis mouse models," *Am. J. Respir. Cell Mol. Biol.*, vol. 36, no. 1, pp. 1–7, 2007, doi: 10.1165/rcmb.2006-0184TR.

[57] R. C. Boucher and B. R. Grubb, "Pathophysiology of gene-targeted mouse models for cystic fibrosis," *Physiol. Rev.*, vol. 79, no. 1 Suppl, pp. S193-214, 1999, [Online]. Available: http://physrev.physiology.org/content/79/1/S193.long.

[58] A. K. Olivier, K. N. Gibson-Corley, and D. K. Meyerholz, "Animal models of gastrointestinal and liver diseases. Animal models of cystic fibrosis: Gastrointestinal, pancreatic, and hepatobiliary disease and pathophysiology," *Am. J. Physiol. - Gastrointest. Liver Physiol.*, vol. 308, no. 6, pp. G459–G471, 2015, doi: 10.1152/ajpgi.00146.2014.

[59] K. L. Tuggle *et al.*, "Characterization of defects in ion transport and tissue development in Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)-knockout rats," *PLoS One*, vol. 9, no. 3, pp. 1–14, 2014, doi: 10.1371/journal.pone.0091253.

[60] K. Jarzabek *et al.*, “Cystic fibrosis as a cause of infertility.,” *Reprod. Biol.*, vol. 4, no. 2, pp. 119–129, 2004.

[61] G. F. Ferrazzano, G. Sangianantoni, T. Cantile, I. Amato, S. Orlando, and A. Ingenito, “Dental enamel defects in Italian children with cystic fibrosis: An observational study,” *Community Dent. Health*, vol. 29, no. 1, pp. 106–109, 2012, doi: 10.1922/CDH_2727Ferrazzano04.

[62] F. S. Xu Jie, Rajagopolan Carthic, Hou Xia, Chen Eugene, Boucher Richard C, “Rabbit models for cystic fibrosis.,” *Pediatr Pulmonol*, pp. 115–93, 2016.

[63] Y. Hasin, M. Seldin, and A. Lusis, “Multi-omics approaches to disease,” *Genome Biol.*, vol. 18, no. 1, pp. 1–15, 2017, doi: 10.1186/s13059-017-1215-1.

[64] E. S. Lander *et al.*, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, pp. 860–921, 2001, doi: 10.1038/35057062.

[65] D. C. Chambers, A. M. Carew, S. W. Lukowski, and J. E. Powell, “Transcriptomics and single-cell RNA-sequencing,” *Respirology*, vol. 24, no. 1, pp. 29–36, 2019, doi: 10.1111/resp.13412.

[66] M. Mann and O. N. Jensen, “Proteomic analysis of post-translational modifications,” *Nat. Biotechnol.*, vol. 21, no. 3, pp. 255–261, 2003, doi: 10.1038/nbt0303-255.

[67] A. Jabeen, A. Mohamedali, and S. Ranganathan, “Looking for Missing Proteins,” *Ref. Modul. Life Sci.*, 2019, doi: 10.1016/b978-0-12-809633-8.20167-2.

[68] N. Liessi, N. Pedemonte, A. Armirotti, and C. Braccia, “Proteomics and metabolomics for cystic fibrosis research,” *Int. J. Mol. Sci.*, vol. 21, no. 15, pp. 1–19, 2020, doi: 10.3390/ijms21155439.

[69] S. D. Patterson and R. H. Aebersold, “Proteomics: the first decade and beyond.,” *Nat. Genet.*, vol. 33 Suppl, no. march, pp. 311–23, 2003, doi: 10.1038/ng1106.

[70] M. Walhout, M. Vidal, and J. Dekker, *Handbook of Systems Biology*. 2013.

[71] J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong, and C. M. Whitehouse, “Electrospray Ionization for Mass Spectrometry of Large Biomolecules,” vol. 246, no. 6.

[72] J. B. Fenn, “Electrospray wings for molecular elephants (Nobel lecture),” *Angew. Chemie - Int. Ed.*, vol. 42, no. 33, pp. 3871–3894, 2003, doi: 10.1002/anie.200300605.

[73] Y. Zhang, B. R. Fonslow, B. Shan, M. C. Baek, and J. R. Yates, “Protein analysis by shotgun/bottom-up proteomics,” *Chem. Rev.*, vol. 113, no. 4, pp. 2343–2394, 2013, doi: 10.1021/cr3003533.

[74] K. Breuker, M. Jin, X. Han, H. Jiang, and F. W. McLafferty, “Top-Down Identification and Characterization of Biomolecules by Mass Spectrometry,” *J. Am. Soc. Mass Spectrom.*, vol. 19, no. 8, pp. 1045–1053, 2008, doi: 10.1016/j.jasms.2008.05.013.

[75] J. C. Tran *et al.*, “Mapping intact protein isoforms in discovery mode using top-down proteomics,” *Nature*, vol. 480, no. 7376, pp. 254–258, 2011, doi: 10.1038/nature10575.

[76] R. Wu *et al.*, “A large-scale method to measure absolute protein phosphorylation stoichiometries,” *Nat. Methods*, vol. 8, no. 8, pp. 677–683, 2011, doi: 10.1038/nmeth.1636.

[77] P. G. Sadowski *et al.*, “Quantitative proteomic approach to study subcellular localization of membrane proteins,” *Nat. Protoc.*, vol. 1, no. 4, pp. 1778–1789, 2006, doi: 10.1038/nprot.2006.254.

[78] N. Nikolovski *et al.*, "Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics," *Plant Physiol.*, vol. 160, no. 2, pp. 1037–1051, 2012, doi: 10.1104/pp.112.204263.

[79] A. Christoforou *et al.*, "A draft map of the mouse pluripotent stem cell spatial proteome," *Nat. Commun.*, vol. 7, 2016, doi: 10.1038/ncomms9992.

[80] D. J. Nightingale, A. Geladaki, L. M. Breckels, S. G. Oliver, and K. S. Lilley, "The subcellular organisation of Saccharomyces cerevisiae," *Curr. Opin. Chem. Biol.*, vol. 48, pp. 86–95, 2019, doi: 10.1016/j.cbpa.2018.10.026.

[81] D. J. H. Nightingale, S. G. Oliver, and K. S. Lilley, "Mapping the Saccharomyces cerevisiae Spatial Proteome with High Resolution Using hyperLOPIT," *Methods Mol. Biol.*, vol. 2049, pp. 165–190, 2019, doi: 10.1007/978-1-4939-9736-7_10.

[82] E. Villanueva *et al.*, "Efficient recovery of the RNA-bound proteome and protein-bound transcriptome using phase separation (OOPS)," *Nat. Protoc.*, vol. 15, no. 8, pp. 2568–2588, 2020, doi: 10.1038/s41596-020-0344-2.

[83] D. Korona *et al.*, "Characterisation of protein isoforms encoded by the drosophila glycogen synthase kinase 3 gene shaggy," *PLoS One*, vol. 15, no. 8 August, 2020, doi: 10.1371/journal.pone.0236679.

[84] E. T. Wang *et al.*, "Dysregulation of mRNA localization and translation in genetic disease," *J. Neurosci.*, vol. 36, no. 45, pp. 11418–11426, 2016, doi: 10.1523/JNEUROSCI.2352-16.2016.

[85] A. Masotti, "Mendelian disorders of membrane trafficking," *N. Engl. J. Med.*, vol. 365, no. 21, pp. 2038–2039, 2011, doi: 10.1056/NEJMc1111685.

[86] V. M. Olkkonen and E. Ikonen, "When intracellular logistics fails - Genetics defects in membrane trafficking," *J. Cell Sci.*, vol. 119, no. 24, pp. 5031–5045, 2006, doi: 10.1242/jcs.03303.

[87] A. Geladaki *et al.*, "Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics," *Nat. Commun.*, vol. 10, no. 1, pp. 1–15, 2019, doi: 10.1038/s41467-018-08191-w.

[88] L. Gatto, L. M. Breckels, S. Wieczorek, T. Burger, and K. S. Lilley, "Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata," *Bioinformatics*, vol. 30, no. 9, pp. 1322–1324, 2014, doi: 10.1093/bioinformatics/btu013.

[89] O. M. Crook, L. M. Breckels, K. S. Lilley, P. D. W. Kirk, and L. Gatto, "Open Peer Review A Bioconductor workflow for the Bayesian analysis of spatial proteomics [version 1; peer review: 1 approved, 2 approved with reservations]," pp. 1–32, 2019, doi: 10.12688/f1000research.18636.1.

[90] K. Hollywood, D. R. Brison, and R. Goodacre, "Metabolomics: Current technologies and future trends," *Proteomics*, vol. 6, no. 17, pp. 4716–4723, 2006, doi: 10.1002/pmic.200600106.

[91] S. C. Gates and C. C. Sweeley, "Quantitative metabolic profiling based on gas chromatography," *Clin. Chem.*, vol. 24, no. 10, pp. 1663–1673, 1978, doi: 10.1093/clinchem/24.10.1663.

[92] W. J. Griffiths and Y. Wang, "Mass spectrometry: From proteomics to metabolomics and lipidomics," *Chem. Soc. Rev.*, vol. 38, no. 7, pp. 1882–1896, 2009, doi: 10.1039/b618553n.

[93] D. J. Beale *et al.*, *Review of recent developments in GC–MS approaches to metabolomics-based research*, vol. 14, no. 11. Springer US, 2018.

[94] C. Jang, L. Chen, and J. D. Rabinowitz, “Metabolomics and Isotope Tracing,” *Cell*, vol. 173, no. 4, pp. 822–837, 2018, doi: 10.1016/j.cell.2018.03.055.

[95] S. Dietmair, N. E. Timmins, P. P. Gray, L. K. Nielsen, and J. O. Krömer, “Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol,” *Anal. Biochem.*, vol. 404, no. 2, pp. 155–164, 2010, doi: 10.1016/j.ab.2010.04.031.

[96] C. L. Winder *et al.*, “Global metabolic profiling of Escherichia coli cultures: An evaluation of methods for quenching and extraction of intracellular metabolites,” *Anal. Chem.*, vol. 80, no. 8, pp. 2939–2948, 2008, doi: 10.1021/ac7023409.

[97] A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod, and J. A. McLean, “Untargeted Metabolomics Strategies—Challenges and Emerging Directions,” *J. Am. Soc. Mass Spectrom.*, vol. 27, no. 12, pp. 1897–1905, 2016, doi: 10.1007/s13361-016-1469-y.

[98] R. M. Salek, C. Steinbeck, M. R. Viant, R. Goodacre, and W. B. Dunn, “The role of reporting standards for metabolite annotation and identification in metabolomic studies,” *Gigascience*, vol. 2, no. 1, pp. 2–4, 2013, doi: 10.1186/2047-217X-2-13.

[99] M. Katajamaa, J. Miettinen, and M. Orešič, “MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data,” *Bioinformatics*, vol. 22, no. 5, pp. 634–636, 2006, doi: 10.1093/bioinformatics/btk039.

[100] C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan, and G. Siuzdak, “XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification,” *Anal. Chem.*, vol. 78, no. 3, pp. 779–787, 2006, doi: 10.1021/ac051437y.

[101] A. Lommen and H. J. Kools, “MetAlign 3.0: Performance enhancement by efficient use of advances in computer hardware,” *Metabolomics*, vol. 8, no. 4, pp. 719–726, 2012, doi: 10.1007/s11306-011-0369-1.

[102] C. A. Smith *et al.*, “METLIN: A metabolite mass spectral database,” *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, 2005, doi: 10.1097/01.ftd.0000179845.53213.39.

[103] D. S. Wishart *et al.*, “HMDB 3.0-The Human Metabolome Database in 2013,” *Nucleic Acids Res.*, vol. 41, no. D1, 2013, doi: 10.1093/nar/gks1065.

[104] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, and O. Yanes, “Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects,” *TrAC - Trends Anal. Chem.*, vol. 78, pp. 23–35, 2016, doi: 10.1016/j.trac.2015.09.005.

[105] L. W. Sumner *et al.*, “Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI),” *Metabolomics*, vol. 3, no. 3, pp. 211–221, 2007, doi: 10.1007/s11306-007-0082-2.

[106] X. Wang *et al.*, “Hsp90 Cochaperone Aha1 Downregulation Rescues Misfolding of CFTR in Cystic Fibrosis,” *Cell*, vol. 127, no. 4, pp. 803–815, 2006, doi: 10.1016/j.cell.2006.09.043.

[107] S. Pankow *et al.*, “Δf508 CFTR interactome remodelling promotes rescue of cystic fibrosis,” *Nature*, vol. 528, no. 7583, pp. 510–516, 2015, doi: 10.1038/nature15729.

[108] J. D. Santos *et al.*, “Folding Status Is Determinant over Traffic-Competence in Defining CFTR Interactors in the Endoplasmic Reticulum,” *Cells*, vol. 8, no. 4, p. 353, 2019, doi:

10.3390/cells8040353.

[109] A. M. Matos, F. R. Pinto, P. Barros, M. D. Amaral, R. Pepperkok, and P. Matos, “Inhibition of calpain 1 restores plasma membrane stability to pharmacologically rescued Phe508del-CFTR variant,” *J. Biol. Chem.*, vol. 294, no. 36, pp. 13396–13410, 2019, doi: 10.1074/jbc.RA119.008738.

[110] S. Pankow, C. Bamberger, and J. R. Yates, “A posttranslational modification code for CFTR maturation is altered in cystic fibrosis,” *Sci. Signal.*, vol. 12, no. 562, 2019, doi: 10.1126/scisignal.aan7984.

[111] T. Frischer, J. K. Myung, G. Maurer, I. Eichler, Z. Szepfalusi, and G. Lubec, “Possible dysregulation of chaperon and metabolic proteins in cystic fibrosis bronchial tissue,” *Proteomics*, vol. 6, no. 11, pp. 3381–3388, 2006, doi: 10.1002/pmic.200500487.

[112] M. Puglia *et al.*, “The proteome speciation of an immortalized cystic fibrosis cell line: New perspectives on the pathophysiology of the disease,” *J. Proteomics*, vol. 170, pp. 28–42, 2018, doi: 10.1016/j.jprot.2017.09.013.

[113] R. N., G. V., B. W.E., and Y. J.R., “Quantitative proteomic profiling reveals differentially regulated proteins in cystic fibrosis cells,” *J. Proteome Res.*, vol. 13, no. 11, pp. 4668–4675, 2014, [Online]. Available: http://pubs.acs.org/journal/jprobs%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed12&NEWS=N&AN=2014895892.

[114] D. R. Wetmore *et al.*, “Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells,” *J. Biol. Chem.*, vol. 285, no. 40, pp. 30516–30522, 2010, doi: 10.1074/jbc.M110.140806.

[115] S. Chin *et al.*, “Cholesterol Interaction Directly Enhances Intrinsic Activity of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR),” *Cells*, vol. 8, no. 8, 2019, doi: 10.3390/cells8080804.

[116] E. Joseloff *et al.*, “Serum metabolomics indicate altered cellular energy metabolism in children with cystic fibrosis,” *Pediatr. Pulmonol.*, vol. 49, no. 5, pp. 463–472, 2014, doi: 10.1002/ppul.22859.

[117] B. T. Kopp *et al.*, “Metabolomic responses to lumacaftor/ivacaftor in cystic fibrosis,” *Pediatr. Pulmonol.*, vol. 53, no. 5, pp. 583–591, 2018, doi: 10.1002/ppul.23972.

[118] V. Ogilvie *et al.*, “Differential global gene expression in cystic fibrosis nasal and bronchial epithelium,” *Genomics*, vol. 98, no. 5, pp. 327–336, 2011, doi: 10.1016/j.ygeno.2011.06.008.

[119] S. Worgall *et al.*, “Similarity of gene expression patterns in human alveolar macrophages in response to Pseudomonas aeruginosa and Burkholderia cepacia,” *Infect. Immun.*, vol. 73, no. 8, pp. 5262–5268, 2005, doi: 10.1128/IAI.73.8.5262-5268.2005.

[120] C. Verhaeghe *et al.*, “Role of IKK and ERK pathways in intrinsic inflammation of cystic fibrosis airways,” *Biochem. Pharmacol.*, vol. 73, no. 12, pp. 1982–1994, 2007, doi: 10.1016/j.bcp.2007.03.019.

[121] C. M. P. Ribeiro *et al.*, “Azithromycin treatment alters gene expression in inflammatory, lipid metabolism, and cell cycle pathways in well-differentiated human airway epithelia,” *PLoS One*, vol. 4, no. 6, 2009, doi: 10.1371/journal.pone.0005806.

[122] T. Sun *et al.*, "Transcriptomic responses to ivacaftor and prediction of ivacaftor clinical responsiveness," *Am. J. Respir. Cell Mol. Biol.*, vol. 61, no. 5, pp. 643–652, 2019, doi: 10.1165/rcmb.2019-0032OC.

[123] B. T. Kopp *et al.*, "Whole-blood transcriptomic responses to lumacaftor/ivacaftor therapy in cystic fibrosis," *J. Cyst. Fibros.*, vol. 19, no. 2, pp. 245–254, 2020, doi: 10.1016/j.jcf.2019.08.021.

[124] K. Jiang *et al.*, "RNA sequencing data from neutrophils of patients with cystic fibrosis reveals potential for developing biomarkers for pulmonary exacerbations," *J. Cyst. Fibros.*, vol. 18, no. 2, pp. 194–202, 2019, doi: 10.1016/j.jcf.2018.05.014.

[125] G. Voisin, G. F. Bouvet, P. Legendre, A. Dagenais, C. Massé, and Y. Berthiaume, "Oxidative stress modulates the expression of genes involved in cell survival in ΔF508 cystic fibrosis airway epithelial cells," *Physiol. Genomics*, vol. 46, no. 17, pp. 634–646, 2014, doi: 10.1152/physiolgenomics.00003.2014.

[126] P. L. Zeitlin *et al.*, "Digitoxin for airway inflammation in cystic fibrosis: Preliminary assessment of safety, pharmacokinetics, and dose finding," *Ann. Am. Thorac. Soc.*, vol. 14, no. 2, pp. 220–229, 2017, doi: 10.1513/AnnalsATS.201608-649OC.

[127] D. Polineni *et al.*, "Airway mucosal host defense is key to genomic regulation of cystic fibrosis lung disease severity," *Am. J. Respir. Crit. Care Med.*, vol. 197, no. 1, pp. 79–93, 2018, doi: 10.1164/rccm.201701-0134OC.

[128] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics.," *Nature*, vol. 422, no. 6928, pp. 198–207, 2003, doi: 10.1038/nature01511.

[129] P. Kebarle and U. H. Verkerk, "Electrospray : From Ions in Solution To Ions in the Gas Phase , What We Know Now," *Mass Spectrom. Rev.*, vol. 28, pp. 898–917, 2009, doi: 10.1002/mas.

[130] K. M. Downard, "1912: A titanic year for mass spectrometry," *J. Mass Spectrom.*, vol. 47, no. 8, pp. 1034–1039, 2012, doi: 10.1002/jms.3071.

[131] R. A. Zubarev and A. Makarov, "Orbitrap mass spectrometry," *Anal. Chem.*, vol. 85, no. 11, pp. 5288–5296, 2013, doi: 10.1021/ac4001223.

[132] L. Krasny *et al.*, "SWATH mass spectrometry as a tool for quantitative profiling of the matrisome," *J. Proteomics*, vol. 189, no. February, pp. 11–22, 2018, doi: 10.1016/j.jprot.2018.02.026.

[133] L. C. Gillet *et al.*, "Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis," *Mol. Cell. Proteomics*, vol. 11, no. 6, pp. 1–17, 2012, doi: 10.1074/mcp.O111.016717.

[134] J. Muntel *et al.*, "Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time," *J. Proteome Res.*, vol. 18, no. 3, pp. 1340–1351, 2019, doi: 10.1021/acs.jproteome.8b00898.

[135] R. S. Plumb *et al.*, "UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation," *Rapid Commun. Mass Spectrom.*, vol. 20, no. 13, pp. 1989–1994, 2006, doi: 10.1002/rcm.2550.

[136] A. L. and H. C.L., "Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins," *Mol. Cell. Proteomics*, vol. 5, no. 4, pp. 573–588, 2006, [Online]. Available:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006169699.

[137] B. K. Erickson *et al.*, "Active Instrument Engagement Combined with a Real-Time Database Search for Improved Performance of Sample Multiplexing Workflows," *J. Proteome Res.*, vol. 18, no. 3, pp. 1299–1306, 2019, doi: 10.1021/acs.jproteome.8b00899.

[138] C. Kumar and M. Mann, "Bioinformatics analysis of mass spectrometry-based proteomics data sets," *FEBS Lett.*, vol. 583, no. 11, pp. 1703–1712, 2009, doi: 10.1016/j.febslet.2009.03.035.

[139] A. Kalderimis, R. Stepan, J. Sullivan, R. Lyne, M. Lyne, and G. Micklem, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.," *F1000Research*, vol. 3, 2014, doi: 10.1109/TSMC.1981.4308636.

[140] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, "The KEGG resource for deciphering the genome," *Nucleic Acids Res.*, vol. 32, no. DATABASE ISS., 2004, doi: 10.1093/nar/gkh063.

[141] J. Chong *et al.*, "MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W486–W494, 2018, doi: 10.1093/nar/gky310.

[142] J. Chaker, E. Gilles, T. Leger, B. Jegou, and A. David, "From metabolomics to HRMS-based exposomics: Adapting peak picking and developing scoring for MS1 suspect screening," *Anal. Chem.*, 2021, doi: 10.1021/acs.analchem.0c04660.

[143] R Developpement Core Team, "R: a language and environment for statistical computing," 2010.

[144] J. P. Lambert *et al.*, "Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition," *Nat. Methods*, vol. 10, no. 12, pp. 1239–1245, 2013, doi: 10.1038/nmeth.2702.

[145] J. P. Clancy, "Rapid therapeutic advances in CFTR modulator science," *Pediatr. Pulmonol.*, vol. 53, no. June, pp. S4–S11, 2018, doi: 10.1002/ppul.24157.

[146] S. K. Willsie, "Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR," *Yearb. Med.*, vol. 2016, pp. 214–216, 2016, doi: 10.1016/S0084-3873(16)30200-0.

[147] National Institutes of Health (NIH), "A Study to Evaluate the Efficacy and Safety of VX-661 in Combination With Ivacaftor in Subjects Aged 12 Years and Older With Cystic Fibrosis, Heterozygous for the F508del-CFTR Mutation. Study ID: NCT02516410," 2018, [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT02516410?term=661-107&draw=2&rank=1.

[148] V. Tomati *et al.*, "Genetic Inhibition of the Ubiquitin Ligase Rnf5 Attenuates Phenotypes Associated to F508del Cystic Fibrosis Mutation," *Sci. Rep.*, vol. 5, no. February, pp. 1–17, 2015, doi: 10.1038/srep12138.

[149] N. Liessi *et al.*, "Distinctive lipid signatures of bronchial epithelial cells associated with cystic fibrosis drugs, including Trikafta," *JCI Insight*, vol. 5, no. 16, 2020, doi: 10.1172/jci.insight.138722.

[150] G. Rosenberger *et al.*, "A repository of assays to quantify 10,000 human proteins by SWATH-MS," *Sci. Data*, vol. 1, pp. 1–15, 2014, doi: 10.1038/sdata.2014.31.

[151] C. Braccia, V. Tomati, E. Caci, N. Pedemonte, and A. Armirotti, "SWATH label-free proteomics for cystic fibrosis research," *J. Cyst. Fibros.*, vol. 18, no. 4, pp. 501–506, 2019, doi: 10.1016/j.jcf.2018.10.004.

[152] M. Wilm *et al.*, “Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry,” *Nature*, vol. 379, no. 6564, pp. 466–469, 1996, doi: 10.1038/379466a0.

[153] S. Li *et al.*, “Digging More Missing Proteins Using an Enrichment Approach with ProteoMiner,” *J. Proteome Res.*, vol. 16, no. 12, pp. 4330–4339, 2017, doi: 10.1021/acs.jproteome.7b00353.

[154] H. Wang, S. Sun, Y. Zhang, S. Chen, P. Liu, and B. Liu, “An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 974, pp. 90–95, 2015, doi: 10.1016/j.jchromb.2014.10.031.

[155] I. V. Shilov *et al.*, “The paragon algorithm, a next generation search engine that uses sequence temperature values sequence temperature values and feature probabilities to identify peptides from tandem mass spectra,” *Mol. Cell. Proteomics*, vol. 6, no. 9, pp. 1638–1655, 2007, doi: 10.1074/mcp.T600050-MCP200.

[156] W. H. Tang, I. V. Shilov, and S. L. Seymour, “Nonlinear fitting method for determining local false discovery rates from decoy database searches,” *J. Proteome Res.*, vol. 7, no. 9, pp. 3661–3667, 2008, doi: 10.1021/pr070492f.

[157] D. Szklarczyk *et al.*, “STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.

[158] R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf, “Centering, scaling, and transformations: Improving the biological information content of metabolomics data,” *BMC Genomics*, vol. 7, pp. 1–15, 2006, doi: 10.1186/1471-2164-7-142.

[159] B. Jassal *et al.*, “The reactome pathway knowledgebase,” *Nucleic Acids Res.*, vol. 48, no. D1, pp. D498–D503, 2020, doi: 10.1093/nar/gkz1031.

[160] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, “GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists,” *BMC Bioinformatics*, vol. 10, 2009, doi: 10.1186/1471-2105-10-48.

[161] A. Stojmirović and Y. K. Yu, “Robust and accurate data enrichment statistics via distribution function of sum of weights,” *Bioinformatics*, vol. 26, no. 21, pp. 2752–2759, 2010, doi: 10.1093/bioinformatics/btq511.

[162] W. Shi *et al.*, “A comprehensive functional analysis of tissue specificity of human gene expression,” *BMC Biol.*, vol. 6, no. 1, 2008, [Online]. Available: http://dx.doi.org/10.1186/1741-7007-6-49.

[163] V. Tomati *et al.*, “High-throughput screening identifies FAU protein as a regulator of mutant cystic fibrosis transmembrane conductance regulator channel,” *J. Biol. Chem.*, vol. 293, no. 4, pp. 1203–1217, 2018, doi: 10.1074/jbc.M117.816595.

[164] Y. Zhen *et al.*, “Nuclear Import of Exogenous FGF1 Requires the ER-Protein LRRC59 and the Importins Kpnα1 and Kpnβ1,” *Traffic*, vol. 13, no. 5, pp. 650–664, 2012, doi: 10.1111/j.1600-0854.2012.01341.x.

[165] X. Han *et al.*, “Destabilizing LSD1 by Jade-2 promotes neurogenesis: An antibraking system in neural development,” *Mol. Cell*, vol. 55, no. 3, pp. 482–494, 2014, doi: 10.1016/j.molcel.2014.06.006.

[166] L. V. J. Galietta, S. Jayaraman, and A. S. Verkman, “Cell-based assay for high-throughput quantitative screening of CFTR chloride transport agonists,” *Am. J. Physiol. - Cell Physiol.*, vol. 281, no. 5 50-5, 2001, doi: 10.1152/ajpcell.2001.281.5.c1734.

[167] N. Fiorini *et al.*, “PubMed Labs: An experimental system for improving biomedical literature search,” *Database*, vol. 2018, no. 2018, 2018, doi: 10.1093/database/bay094.

[168] Z. Lu, “PubMed and beyond: A survey of web tools for searching biomedical literature,” *Database*, vol. 2011, 2011, doi: 10.1093/database/baq036.

[169] D. Mendez *et al.*, “ChEMBL: Towards direct deposition of bioassay data,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D930–D940, 2019, doi: 10.1093/nar/gky1075.

[170] W. DS *et al.*, “DrugBank: a comprehensive resource for in silico drug discovery and exploration,” *Nucleic Acids Res.*, vol. 1, no. 34, pp. 668–672, 2006.

[171] C. B. Merrill *et al.*, “Patch clamp-assisted single neuron lipidomics,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, 2017, doi: 10.1038/s41598-017-05607-3.

[172] S. Wickramasekara *et al.*, “Electrospray Quadrupole Travelling Wave Ion Mobility Time-of-Flight Mass Spectrometry for the Detection of Plasma Metabolome Changes Caused by Xanthohumol in Obese Zucker (fa/fa) Rats,” *Metabolites*, vol. 3, no. 3, pp. 701–717, 2013, doi: 10.3390/metabo3030701.

[173] A. Basit, D. Piomelli, and A. Armirotti, “Rapid evaluation of 25 key sphingolipids and phosphosphingolipids in human plasma by LC-MS/MS,” *Anal. Bioanal. Chem.*, vol. 407, no. 17, pp. 5189–5198, 2017, doi: 10.1007/s00216-015-8585-6.

[174] E. Szymańska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, “Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies,” *Metabolomics*, vol. 8, pp. 3–16, 2012, doi: 10.1007/s11306-011-0330-3.

[175] S. Mondal and H. Mondal, “Value of r2 in statistical analysis by Pearson correlation coefficient,” *J. Clin. Diagnostic Res.*, vol. 11, no. 11, p. CL01, 2017, doi: 10.7860/JCDR/2017/29763.10812.

[176] D. S. Wishart *et al.*, “HMDB 4.0: The human metabolome database for 2018,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D608–D617, 2018, doi: 10.1093/nar/gkx1089.

[177] A. Marco-Ramell *et al.*, “Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data,” *BMC Bioinformatics*, vol. 19, no. 1, pp. 1–11, 2018, doi: 10.1186/s12859-017-2006-0.

[178] E. Fahy, M. Sud, D. Cotter, and S. Subramaniam, “LIPID MAPS online tools for lipid research,” *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 606–612, 2007, doi: 10.1093/nar/gkm324.

[179] T. Cajka and O. Fiehn, “Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry,” *TrAC - Trends Anal. Chem.*, vol. 61, pp. 192–206, 2014, doi: 10.1016/j.trac.2014.04.017.

[180] R. C. Gentleman *et al.*, “Bioconductor: open software development for computational biology and bioinformatics.,” *Genome Biol.*, vol. 5, no. 10, 2004, doi: 10.1186/gb-2004-5-10-r80.

[181] L. Gatto, L. M. Breckels, C. M. Mulvey, and K. S. Lilley, “A Bioconductor workflow for processing and analysing spatial proteomics data [version 2; referees: 2 approved],” *F1000Research*, vol. 5,

2018, doi: 10.12688/f1000research.10411.2.

[182] M. Yano, Y. Koumoto, Y. Kanesaki, X. Wu, and H. Kido, “20S proteasome prevents aggregation of heat-denatured protein without PA700 regulatory subcomplex like a molecular chaperone,” *Biomacromolecules*, vol. 5, no. 4, pp. 1465–1469, 2004, doi: 10.1021/bm049957a.

[183] J. Moelleken *et al.*, “Differential localization of coatomer complex isoforms within the Golgi apparatus,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 11, pp. 4425–4430, 2007, doi: 10.1073/pnas.0611360104.

[184] M. L. Styers, A. K. O’Connor, R. Grabski, E. Cormet-Boyaka, and E. Sztul, “Depletion of β-COP reveals a role for COP-I in compartmentalization of secretory compartments and in biosynthetic transport of caveolin-1,” *Am. J. Physiol. - Cell Physiol.*, vol. 294, no. 6, 2008, doi: 10.1152/ajpcell.00010.2008.

[185] S. Alberti, K. Böhse, V. Arndt, A. Schmitz, and J. Höhfeld, “The cochaperone HspBP1 inhibits the CHIP ubiquitin ligase and stimulates the maturation of the cystic fibrosis transmembrane conductance regulator,” *Mol. Biol. Cell*, vol. 15, no. 9, pp. 4003–4010, 2004, doi: 10.1091/mbc.E04-04-0293.

[186] J. Rodriguez *et al.*, “Substrate-Trapped Interactors of PHD3 and FIH Cluster in Distinct Signaling Pathways,” *Cell Rep.*, vol. 14, no. 11, pp. 2745–2760, 2016, doi: 10.1016/j.celrep.2016.02.043.

[187] M. Kanehisa and S. Goto, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000, doi: 10.1093/nar/28.1.27.

[188] T. Zhu, D. Dahan, A. Evagelidis, S. X. Zheng, J. Luo, and J. W. Hanrahan, “Association of cystic fibrosis transmembrane conductance regulator and protein phosphatase 2C,” *J. Biol. Chem.*, vol. 274, no. 41, pp. 29102–29107, 1999, doi: 10.1074/jbc.274.41.29102.

[189] J. P. Garnett, E. Hickman, O. Tunkamnerdthai, A. W. Cuthbert, and M. A. Gray, “Protein phosphatase 1 coordinates CFTR-dependent airway epithelial HCO 3- secretion by reciprocal regulation of apical and basolateral membrane Cl--HCO3-exchangers,” *Br. J. Pharmacol.*, vol. 168, no. 8, pp. 1946–1960, 2013, doi: 10.1111/bph.12085.

[190] M. E. Price *et al.*, “S-nitrosation of protein phosphatase 1 mediates alcohol-induced ciliary dysfunction,” *Sci. Rep.*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-27924-x.

[191] K. Zaman *et al.*, “S-nitrosothiols increases cystic fibrosis transmembrane regulator expression and maturation in the cell surface,” *Biochem. Biophys. Res. Commun.*, vol. 443, no. 4, pp. 1257–1262, 2014, doi: 10.1016/j.bbrc.2013.12.130.

[192] M. Rossignol, A. Keriel, A. Staub, and J. M. Egly, “Kinase activity and phosphorylation of the largest subunit of TFIIF transcription factor,” *J. Biol. Chem.*, vol. 274, no. 32, pp. 22387–22392, 1999, doi: 10.1074/jbc.274.32.22387.

[193] J. J. Oh, D. R. Grosshans, S. G. Wong, and D. J. Slamon, “Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells,” *Nucleic Acids Res.*, vol. 27, no. 20, pp. 4008–4017, 1999, doi: 10.1093/nar/27.20.4008.

[194] Q. Lu *et al.*, “Early steps in primary cilium assembly require EHD1/EHD3-dependent ciliary vesicle formation,” *Nat. Cell Biol.*, vol. 17, no. 3, pp. 228–240, 2015, doi: 10.1038/ncb3109.

[195] G. Euskirchen, R. K. Auerbach, and M. Snyder, “SWI/SNF chromatin-remodeling factors: Multiscale

analyses and diverse functions," *J. Biol. Chem.*, vol. 287, no. 37, pp. 30897–30905, 2012, doi: 10.1074/jbc.R111.309302.

[196] N. Agrawal and R. Banerjee, "Human polycomb 2 protein is a SUMO E3 ligase and alleviates substrate-induced inhibition of cystathionine β-synthase sumoylation," *PLoS One*, vol. 3, no. 12, 2008, doi: 10.1371/journal.pone.0004032.

[197] M. C. De Marco *et al.*, "MAL2, a novel raft protein of the MAL family, is an essential component of the machinery for transcytosis in hepatoma HepG2 cells," *J. Cell Biol.*, vol. 159, no. 1, pp. 37–44, 2002, doi: 10.1083/jcb.200206033.

[198] D. Guan, D. Factor, Y. Liu, Z. Wang, and H. Y. Kao, "The epigenetic regulator UHRF1 promotes ubiquitination-mediated degradation of the tumor-suppressor protein promyelocytic leukemia protein," *Oncogene*, vol. 32, no. 33, pp. 3819–3828, 2013, doi: 10.1038/onc.2012.406.

[199] Y. G. Alevy *et al.*, "IL-13-induced airway mucus production is attenuated by MAPK13 inhibition," *J. Clin. Invest.*, vol. 122, no. 12, pp. 4555–4568, 2012, doi: 10.1172/JCI64896.

[200] L. G. Johnson, J. C. Olsen, B. Sarkadi, K. L. Moore, R. Swanstrom, and R. C. Boucher, "Efficiency of gene transfer for restoration of normal airway epithelial function in cystic fibrosis," *Nat. Genet.*, vol. 2, no. 1, pp. 21–25, 1992, doi: 10.1038/ng0992-21.

[201] H. Caohuy *et al.*, "Activation of 3-phosphoinositide-dependent kinase 1 (PDK1) and serum- And glucocorticoid-induced protein kinase 1 (SGK1) by short-chain sphingolipid C4-ceramide rescues the trafficking defect of &x25b5;F508-cystic fibrosis transmembrane conductance regula," *J. Biol. Chem.*, vol. 289, no. 52, pp. 35953–35968, 2014, doi: 10.1074/jbc.M114.598649.

[202] A. Abu-Arish, E. Pandžić, D. Kim, H. W. Tseng, P. W. Wiseman, and J. W. Hanrahan, "Agonists that stimulate secretion promote the recruitment of CFTR into membrane lipid microdomains," *J. Gen. Physiol.*, vol. 151, no. 6, pp. 834–849, 2018, doi: 10.1085/JGP.201812143.

[203] E. Hildebrandt, N. Khazanov, J. C. Kappes, Q. Dai, H. Senderowitz, and I. L. Urbatsch, "Specific stabilization of CFTR by phosphatidylserine," *Biochim. Biophys. Acta - Biomembr.*, vol. 1859, no. 2, pp. 289–293, 2017, doi: 10.1016/j.bbamem.2016.11.013.

[204] V. A. van der Mark *et al.*, "The phospholipid flippase ATP8B1 mediates apical localization of the cystic fibrosis transmembrane regulator," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1863, no. 9, pp. 2280–2288, 2016, doi: 10.1016/j.bbamcr.2016.06.005.

[205] M. Veltman *et al.*, "Correction of lung inflammation in a F508del CFTR murine cystic fibrosis model by the sphingosine-1-phosphate lyase inhibitor LX2931," *Am. J. Physiol. - Lung Cell. Mol. Physiol.*, vol. 311, no. 5, pp. L1000–L1014, 2016, doi: 10.1152/ajplung.00298.2016.

[206] F. A. Malik *et al.*, "Sphingosine-1-phosphate is a novel regulator of cystic fibrosis transmembrane conductance regulator (CFTR) activity," *PLoS One*, vol. 10, no. 6, 2015, doi: 10.1371/journal.pone.0130313.

[207] D. M. Cholon *et al.*, "Cystic fibrosis: Potentiator ivacaftor abrogates pharmacological correction of ΔF508 CFTR in cystic fibrosis," *Sci. Transl. Med.*, vol. 6, no. 246, pp. 2145–2156, 2014, doi: 10.1126/scitranslmed.3008680.

[208] G. Veit *et al.*, "Cystic fibrosis: Some gating potentiators, including VX-770, diminish ΔF508-CFTR functional expression," *Sci. Transl. Med.*, vol. 6, no. 246, p. 246ra97, 2014, doi:

10.1126/scitranslmed.3008889.

[209] S. Furse and A. I. P. M. De Kroon, “Phosphatidylcholines functions beyond that of a membrane brick,” *Mol. Membr. Biol.*, vol. 32, no. 4, pp. 117–119, 2015, doi: 10.3109/09687688.2015.1066894.

[210] M. Bhagat and S. Sofou, “Membrane heterogeneities and fusogenicity in phosphatidylcholine-phosphatidic acid rigid vesicles as a function of pH and lipid chain mismatch,” *Langmuir*, vol. 26, no. 3, pp. 1666–1673, 2010, doi: 10.1021/la9026283.

[211] M. M. Ulane, J. D. Butler, A. Peri, L. Miele, R. E. Ulane, and V. S. Hubbard, “Cystic fibrosis and phosphatidylcholine biosynthesis,” *Clin. Chim. Acta*, vol. 230, no. 2, pp. 109–116, 1994, doi: 10.1016/0009-8981(94)90263-1.

[212] S. Sahu and W. S. Lynn, “Lipid composition of airway secretions from patients with asthma and patients with cystic fibrosis,” *Am. Rev. Respir. Dis.*, vol. 115, no. 2, pp. 233–239, 1977, doi: 10.1164/arrd.1977.115.2.233.

[213] S. Chin *et al.*, “Lipophilicity of the cystic fibrosis drug, ivacaftor (VX-770), and its destabilizing effect on the major CF-causing mutation: F508del s,” *Mol. Pharmacol.*, vol. 94, no. 2, pp. 917–925, 2018, doi: 10.1124/mol.118.112177.

[214] J. L. Delaunay *et al.*, “Functional defect of variants in the adenosine triphosphate–binding sites of ABCB4 and their rescue by the cystic fibrosis transmembrane conductance regulator potentiator, ivacaftor (VX-770),” *Hepatology*, vol. 65, no. 2, pp. 560–570, 2017, doi: 10.1002/hep.28929.

[215] K. A. Becker *et al.*, “Sphingolipids as targets for inhalation treatment of cystic fibrosis,” *Adv. Drug Deliv. Rev.*, vol. 133, pp. 66–75, 2018, doi: 10.1016/j.addr.2018.04.015.

[216] B. M. Castro, M. Prieto, and L. C. Silva, “Ceramide: A simple sphingolipid with unique biophysical properties,” *Prog. Lipid Res.*, vol. 54, no. 1, pp. 53–67, 2014, doi: 10.1016/j.plipres.2014.01.004.

[217] A. Basit, D. Piomelli, and A. Armirotti, “Rapid evaluation of 25 key sphingolipids and phosphosphingolipids in human plasma by LC-MS/MS,” *Anal. Bioanal. Chem.*, vol. 407, no. 17, pp. 5189–5198, 2017, doi: 10.1007/s00216-015-8585-6.

[218] C. Chan and T. Goldkorn, “Ceramide path in human lung cell death,” *Am. J. Respir. Cell Mol. Biol.*, vol. 22, no. 4, pp. 460–468, 2000, doi: 10.1165/ajrcmb.22.4.3376.

[219] V. Gagliostro *et al.*, “Dihydroceramide delays cell cycle G1/S transition via activation of ER stress and induction of autophagy,” *Int. J. Biochem. Cell Biol.*, vol. 44, no. 12, pp. 2135–2143, 2012, doi: 10.1016/j.biocel.2012.08.025.

[220] K. Lauber *et al.*, “Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal,” *Cell*, vol. 113, no. 6, pp. 717–730, 2003, doi: 10.1016/S0092-8674(03)00422-7.

[221] A. Yamashita, T. Sugiura, and K. Waku, “Acyltransferases and transacylases involved in fatty acid remodeling of phospholipids and metabolism of bioactive lipids in mammalian cells,” *J. Biochem.*, vol. 122, no. 1, pp. 1–16, 1997, doi: 10.1093/oxfordjournals.jbchem.a021715.

[222] O. M. Crook, T. Smith, M. Elzek, and K. S. Lilley, “Moving Profiling Spatial Proteomics Beyond Discrete Classification,” *Proteomics*, 2020, doi: 10.1002/pmic.201900392.

[223] W. J. Chung *et al.*, “Increasing the endoplasmic reticulum pool of the F508del allele of the cystic

fibrosis transmembrane conductance regulator leads to greater folding correction by small molecule therapeutics," *PLoS One*, vol. 11, no. 10, 2016, doi: 10.1371/journal.pone.0163615.

[224] N. Rauniyar, V. Gupta, W. E. Balch, and J. R. Yates, "Quantitative proteomic profiling reveals differentially regulated proteins in cystic fibrosis cells," *J. Proteome Res.*, vol. 13, no. 11, pp. 4668–4675, 2014, doi: 10.1021/pr500370g.

[225] P. W. Phuan *et al.*, "Synergy-based small-molecule screen using a human lung epithelial cell line yields Δf508-CFTR correctors that augment VX-809 maximal efficacy," *Mol. Pharmacol.*, vol. 86, no. 1, pp. 42–51, 2014, doi: 10.1124/mol.114.092478.

[226] N. Pedemonte, V. Tomati, E. Sondo, and L. J. V. Galietta, "Influence of cell background on pharmacological rescue of mutant CFTR," *Am. J. Physiol. - Cell Physiol.*, vol. 298, no. 4, 2010, doi: 10.1152/ajpcell.00404.2009.

[227] K. Simons and E. Ikonen, "Functional rafts in cell membranes," *Nature*, vol. 387, no. 6633, pp. 569–572, 1997, doi: 10.1038/42408.

[228] H. Grassmé *et al.*, "Host defense against Pseudomonas aeruginosa requires ceramide-rich membrane rafts," *Nat. Med.*, vol. 9, no. 3, pp. 322–330, 2003, doi: 10.1038/nm823.

[229] M. P. Kowalski and G. B. Pier, " Localization of Cystic Fibrosis Transmembrane Conductance Regulator to Lipid Rafts of Epithelial Cells Is Required for Pseudomonas aeruginosa -Induced Cellular Activation ," *J. Immunol.*, vol. 172, no. 1, pp. 418–425, 2004, doi: 10.4049/jimmunol.172.1.418.